CONTENTS

MECHANICS

		PAGE
CHAPTER 1	**ROTATIONAL MOTION**	**1**
1.1	ROTATIONAL MOTION UNDER CONSTANT ANGULAR ACCELERATION	1
1.2	RELATIONS BETWEEN LINEAR AND ANGULAR QUANTITIES	4
1.3	CENTRIPETAL ACCELERATION	6
	SUMMARY	11
	EXERCISES	11
	CONCEPT MAP	12
CHAPTER 2	**FORCES IN CIRCULAR MOTION**	**13**
2.1	FORCES CAUSING CENTRIPETAL ACCELERATION	13
2.2	CENTRIPETAL AND CENTRIFUGAL FORCES	18
	SUMMARY	21
	EXERCISES	21
	CONCEPT MAP	22
CHAPTER 3	**FLUID DYNAMICS**	**23**
3.1	LAMINAR AND TURBULENT FLOW	23
3.2	EQUATION OF CONTINUITY FOR FLUIDS	24
3.3	BERNOULLI'S EQUATION	27
3.4	MANIFESTATIONS OF BERNOULLI'S THEOREM	31
3.5	VISCOSITY (Fluid Friction)	33
3.6	SURFACE TENSION AND CAPILLARITY	36
	SUMMARY	39
	EXERCISES	40
	CONCEPT MAP	41

ENERGY

			PAGE
CHAPTER 4	**SOURCES OF ENERGY AND ENVIRONMENTAL IMPACTS**		**42**
	4.1	ENERGY DEMAND AND CRISIS	42
	4.2	SOURCES OF ENERGY	43
	4.3	NON-RENEWABLE SOURCES OF ENERGY	44
	4.4	RENEWABLE SOURCES OF ENERGY	45
	4.5	ENVIRONMENTAL IMPACTS AND CONSEQUENCES	57
		SUMMARY	60
		EXERCISES	61
		CONCEPT MAP	62
CHAPTER 5	**HEAT TRANSFER AND THERMODYNAMICS**		**63**
	5.1	HEAT TRANSFER	63
	5.2	THERMODYNAMICS	71
	5.3	HEAT ENGINES	76
		SUMMARY	82
		EXERCISES	82
		CONCEPT MAP	84
CHAPTER 6	**EFFECTS AND CHARACTERISTICS OF SOUND**		**85**
	6.1	PROPAGATION OF SOUND IN A MEDIUM	85
	6.2	ECHO	88
	6.3	AUDIBILITY	90
	6.4	THE DOPPLER EFFECT	92
	6.5	PROPERTIES OF SOUND WAVES	95
		SUMMARY	97
		EXERCISES	97
		CONCEPT MAP	98

PAGE

CHAPTER 7 APPLICATIONS OF LIGHT AND OPTICAL INSTRUMENTS 99

7.1 THE CAMERA 99
7.2 HUMAN EYE 101
7.3 COMPOUND MICROSCOPE 104
7.4 TELESCOPE 107
7.5 LASER 111
7.6 FIBRE OPTICS 113
7.7 SPECTROMETER 115
7.8 PHOTOMETRY 117
SUMMARY 120
EXERCISES 120
CONCEPT MAP 122

CHAPTER 8 INTERFERENCE AND DIFFRACTION OF LIGHT 123

8.1 HUYGENS' PRINCIPLE 123
8.2 INTERFERENCE OF LIGHT 125
8.3 DIFFRACTION OF LIGHT 130
8.4 DIFFRACTION GRATING 133
SUMMARY 135
EXERCISES 135
CONCEPT MAP 136

ELECTRICITY & MAGNETISM

CHAPTER 9 CAPACITOR AND CAPACİTANCE 137

9.1 CAPACITORS 137
9.2 CAPACITANCE OF A CAPACITOR 139
9.3 PARALLEL-PLATE CAPACITOR 141
9.4 ENERGY STORED IN A CAPACITOR 145
9.5 COMBINATION OF CAPACITORS 148
SUMMARY 151
EXERCISES 151
CONCEPT MAP 152

PAGE

CHAPTER 10 ELECTRICAL ENERGY, POWER AND HEATING EFFECT OF ELECTRIC CURRENT 153

10.1 ELECTROMOTIVE FORCE AND ELECTRIC CIRCUITS 153
10.2 BATTERIES IN SERIES AND IN PARALLEL 158
10.3 ELECTRICAL ENERGY AND ELECTRICAL POWER 160
10.4 JOULE'S LAW OF ELECTRICITY AND HEAT 163
10.5 SOME APPLICATIONS OF THE HEATING EFFECT OF CURRENT 166
SUMMARY 168
EXERCISES 169
CONCEPT MAP 171

CHAPTER 11 ELECTROMAGNETIC INDUCTION, GENERATION AND DISTRIBUTION OF ELECTRICITY 172

11.1 ALTERNATING CURRENT 172
11.2 ELECTROMAGNETIC INDUCTION 175
11.3 FARADAY'S LAW AND LENZ'S LAW 178
11.4 APPLICATIONS OF ELECTROMAGNETIC INDUCTION 180
11.5 POWER TRANSMISSION 186
11.6 PRINCIPLES OF HOUSE WIRING 187
11.7 DANGERS OF ELECTRICITY 190
SUMMARY 193
EXERCISES 193
CONCEPT MAP 195

MODERN PHYSICS

		PAGE
CHAPTER 12	**DIGITAL ELECTRONICS AND COMMUNICATION SYSTEM**	**196**
	12.1 ELECTRONIC SYSTEM	196
	12.2 DIGITAL ELECTRONICS	205
	12.3 BASIC ELECTRONIC COMMUNICATION	212
	SUMMARY	219
	EXERCISES	220
	CONCEPT MAP	222
CHAPTER 13	**MODERN PHYSICS**	**223**
	13.1 RADIOACTIVITY AND USES OF RADIOACTIVITY	223
	13.2 NUCLEAR ENERGY AND ITS ENVIRONMENTAL IMPACT	233
	13.3 WAVE-PARTICLE DUALISM	237
	13.4 PRINCIPLE OF SPECIAL THEORY OF RELATIVITY	239
	SUMMARY	246
	EXERCISES	247
	CONCEPT MAP	248
REFERENCES		249

CHAPTER 1

ROTATIONAL MOTION

In Grade 11, the angular displacement, angular velocity and angular acceleration of an object in circular motion have been learnt. In this chapter, the concepts and equations for rotational motion under constant angular acceleration and centripetal acceleration will be studied.

Learning Outcomes

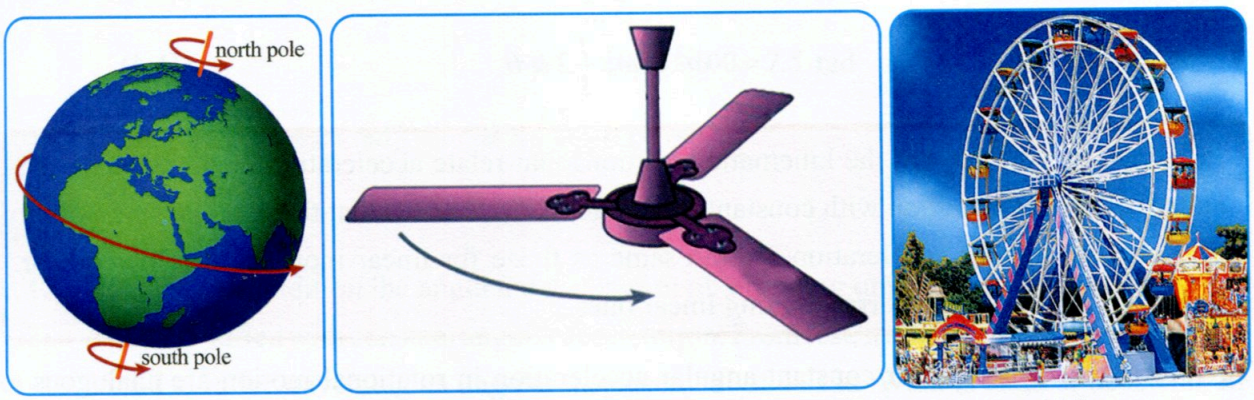
It is expected that students will

- differentiate between the circular motion and rotational motion.
- examine rotational motion under constant angular acceleration.
- deduce the relations between angular and linear quantities.
- understand the centripetal acceleration.
- distinguish between the uniform circular motion and nonuniform circular motion.

1.1 ROTATIONAL MOTION UNDER CONSTANT ANGULAR ACCELERATION

In a circular motion, the object just moves in a circle. For example, artificial satellites going around the earth at a constant height.

In rotational motion, the object rotates about an axis. Some of the examples of rotational motion are; rotation of earth about its own axis which creates the cycle of day and night, motion of the blades of a fan, and motion of a Ferris wheel in an amusement park as shown in Figure 1.1 (a), (b) and (c).



(a) (b) (c)

Figure 1.1 Some examples of rotational motion

A rigid body is a body that does not deform or change shape. That is, no matter how the body moves, the distance between any two particles within the body remains constant. If a rigid body is

in rotational motion, all the particles constituting it undergo circular motion about a common axis. When an object (rigid body) rotates, it can speed up (or) slow down. During these time intervals, its angular velocity is changing therefore it has an angular acceleration. If its angular velocity changes at a costant rate, then we can say that the angular acceleration is constant and the motion is called the rotational motion under constant angular acceleration.

Rotational variables; such as angular displacement, angular velocity, and angular acceleration have been introduced previously. In this section, these variables are applied to analyze the rotational motion for a rigid body about a fixed axis under a constant angular acceleration.

For an object that rotates with a constant angular acceleration, the kinematic equations for rotational motion can be derived in terms of the angular displacement, angular velocity and angular acceleration of the object at any instant of time.

If the initial angular speed (magnitude of angular velocity of the object at $t = 0$) is ω_0, its angular speed ω at time t can be derived from the following constant angular acceleration.

$$\alpha = \frac{\omega - \omega_0}{t}$$

We get, $$\omega = \omega_0 + \alpha t \qquad (1.1)$$

Average angular speed is, $$\overline{\omega} = \frac{\omega + \omega_0}{2} \qquad (1.2)$$

Angular displacement for time interval t is $\theta = \overline{\omega} t$. Using Eq. (1.1) and Eq. (1.2) angular displacement θ and angular speed ω can be expressed by the following equations.

$$\theta = \omega_0 t + \frac{1}{2}\alpha t^2 \qquad (1.3)$$

$$\omega^2 = \omega_0^2 + 2\alpha\theta \qquad (1.4)$$

In Grade 10, we derived the kinematic equations that relate acceleration, velocity, distance, and time for linear motion with constant acceleration. We can see that the equations of motion for constant angular acceleration are the same as those for linear motion, with the angular quantities replaced with corresponding linear ones.

Therefore, the equations for constant angular acceleration in rotational motion are analogous to equations for constant acceleration in linear motion as shown in Table 1.1.

Table 1.1 Analogies between kinematic equations of linear and angular motion

Linear motion	Angular motion
$v = v_0 + a\,t$	$\omega = \omega_0 + \alpha\,t$
$s = v_0 t + \frac{1}{2} a\,t^2$	$\theta = \omega_0 t + \frac{1}{2}\alpha\,t^2$
$v^2 = v_0^2 + 2\,a\,s$	$\omega^2 = \omega_0^2 + 2\,\alpha\,\theta$
$\bar{v} = \frac{v + v_0}{2}$	$\bar{\omega} = \frac{\omega + \omega_0}{2}$

Example 1.1 The Ferris wheel starts from rest and reaches an angular velocity of 1.5 rad s^{-1} over a 10 s period under the constant angular acceleration. (i) Find the angular acceleration of Ferris wheel. (ii) How many revolutions does it make during 10 s?

Initial angular velocity $\omega_0 = 0$, angular velocity ω = 1.5 rad s^{-1}, t = 10 s

(i) The angular acceleration of Ferris wheel can be calculated by using equations of motion for constant angular acceleration.

$$\omega = \omega_0 + \alpha\,t$$
$$1.5 = 0 + \alpha \times 10$$
$$\alpha = 0.15 \text{ rad } s^{-2}$$

(ii) To get number of revolutions in 10 s, first we must find angular displacement.

$$\theta = \omega_0\,t + \frac{1}{2}\alpha\,t^2$$
$$= 0 + \frac{1}{2} \times 0.15 \times 100 = 7.5 \text{ rad}$$

Since , 1 rev = 2π rad, $\theta = \frac{7.5}{2\pi} = 1.19$ rev

Reviewed Exercise

1. Under what condition the angular acceleration of a rotating object is equal to zero?
2. A ball is whirled with constant angular acceleration. From rest, it attains an angular velocity of 25 rad s^{-1} after traversing an angular displacement of 41 rad. What is angular acceleration of the ball?

Key Words: rotational motion, angular displacement, angular velocity, angular acceleration

1.2 RELATIONS BETWEEN LINEAR AND ANGULAR QUANTITIES

The relations between linear quantities (s, v, a) and angular quantities (θ, ω, α) for circular motion have been studied in Grade 11. In this section, the relation between these quantities for rotational motion will be studied. All particles of an object rotating around the axis of rotation move in a circular path.

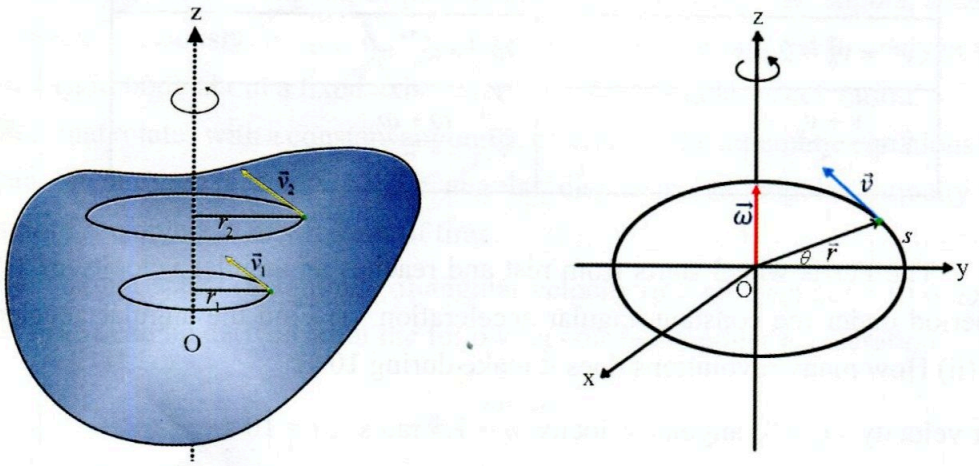


Figure 1.2 (a) Rotational motion of an object (b) circular motion of a particle of the object

Let us consider a randomly shaped rigid body undergoing a rotational motion as shown in Figure 1.2 (a). All particles of the body move in a circle lying on a plane that is perpendicular to the axis, such that the centre of rotation lies on the axis as shown in Figure 1.2 (b).

If the linear distance of a particle of rotating object measured along the arc is s and the arc subtends angle θ at the origin, the relation between arc length s and angular displacement θ, measured in radian, is $s = r\,\theta$ where r is the radius of the circle.

The instantaneous angular velocity ω of rotational object is the same for all the particles comprising the object. The relation between the magnitude of the linear velocity v and the angular velocity ω of a particle is $v = r\,\omega$, where ω is measured in radian per second. This relation applies to every particle of the rotational object. Since the direction of linear velocity is tangent to the path, the linear velocity is also called tangential velocity.

If the magnitude of the angular velocity of a rotating object changes, the magnitude of linear velocity will also change and the object will have a tangential acceleration. The direction of tangential acceleration is always tangent to the circular path. The tangential acceleration a_T relates to angular acceleration α (measured in radian per second squared) by the equation $a_T = r\alpha$.

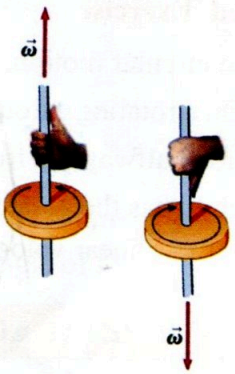
The tangential acceleration of a rotating object is a measure of how fast the tangential velocity changes.

For the rotational motion, Table 1.2 expresses the relation between the linear and rotational quantities.

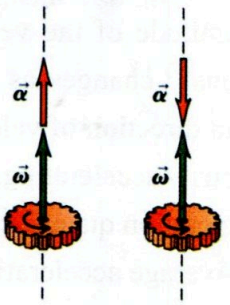
Table 1.2 The relation between the linear and rotational quantities

Quantity	Linear quantities	Rotational quantities	Relation
displacement	s	θ	$s = r\theta$
tangential velocity	v	ω	$v = r\omega$
tangential acceleration	a_T	α	$a_T = r\alpha$

The direction of the angular velocity vector is perpendicular to the plane of rotation, and along the axis of rotation.
The direction of angular velocity is usually specified by the right-hand rule. Wrap the right hand around the axis of rotation so that the fingers are pointing in the direction of rotation. The thumb points in the direction of the angular velocity vector.



When the angular velocity is increasing, the angular acceleration vector points in the same direction as the angular velocity. When the angular velocity is decreasing, the angular acceleration vector points in the opposite to the angular velocity.



Example 1.2 A boy steps on a merry-go-round which has a radius of 5 m and is at rest. It starts accelerating at a constant rate up to an angular velocity of 5 rad s^{-1} in 20 s. What is the distance traversed by the boy?

Radius of merry-go-round $r = 5$ m, initial angular velocity $\omega_0 = 0$,
angular velocity in 20 s, $\omega = 5$ rad s^{-1}, $t = 20$ s
The angular displacement θ that the boy passes through is,

$$\theta = \overline{\omega}\, t = \left(\frac{\omega + \omega_0}{2}\right) t$$

$$= \left(\frac{5 + 0}{2}\right) \times 20 = 50 \text{ rad}$$

The distance traversed by the boy is,

$$s = r\theta$$

$$= 5 \times 50 = 250 \text{ m}$$

Example 1.3 The wheel of a car with radius 20 cm starts moving. The angular acceleration provided by the engine is 12 rad s^{-2}. What is the tangential acceleration of rim of the wheel?

Radius of wheel $r = 20$ cm $= 0.2$ m, angular acceleration $\alpha = 12$ rad s^{-2}

The tangential acceleration of rim of the wheel is,

$$a_T = r\alpha$$
$$= 0.2 \times 12 = 2.4 \text{ m s}^{-2}$$

The direction of tangential acceleration is tangent to rim of the wheel.

Reviewed Exercise

1. In circular motion, how is a tangential acceleration produced?
2. On a rotating carousel, a child sits on a horse near the outer edge and another child sits on a lion halfway out from the centre. (i) Which child has the greater linear velocity? (ii) Which child has the greater angular velocity?

Key Words: linear velocity, tangential acceleration

1.3 CENTRIPETAL ACCELERATION

An object that moves in a circle at constant speed v is said to experience uniform circular motion. The magnitude of the velocity remains constant in this case, but the direction of the velocity continuously changes as the object moves around the circle as shown in Figure 1.3 (a). The change in direction of velocity constitutes an acceleration. Thus, an object revolving in a circle is continuously accelerating, even when the speed remains constant ($v_1 = v_2 = v$). We now investigate this acceleration quantitatively.

Average acceleration has been defined as

$$\vec{a} = \frac{\vec{v}_2 - \vec{v}_1}{\Delta t} = \frac{\Delta \vec{v}}{\Delta t} \qquad (1.5)$$

where $\Delta \vec{v}$ is the change in velocity during the time interval Δt. Consider the situation in which Δt approaches zero and thus the instantaneous acceleration can be obtained. During the time interval Δt, the object moves from point A to point B, covering a distance Δs along the arc which subtends an angle $\Delta \theta$. The change in the velocity vector is $\vec{v}_2 - \vec{v}_1 = \Delta \vec{v}$, and is shown in Figure 1.3 (b).

If Δt is very small (approaching zero), then Δs and $\Delta \theta$ are also very small, and, $\vec{v}_2$ will be almost parallel to $\vec{v}_1$. $\Delta \vec{v}$ will be essentially perpendicular to both of them as shown in Figure 1.3 (c) and points toward the centre of the circle. Acceleration $\vec{a}$ is in the same direction as $\Delta \vec{v}$, it also points toward the centre of the circle. Therefore, this acceleration is centripetal acceleration or radial acceleration (since it is directed along the radius towards the centre of the circle), and denoted by $\vec{a}_C$.

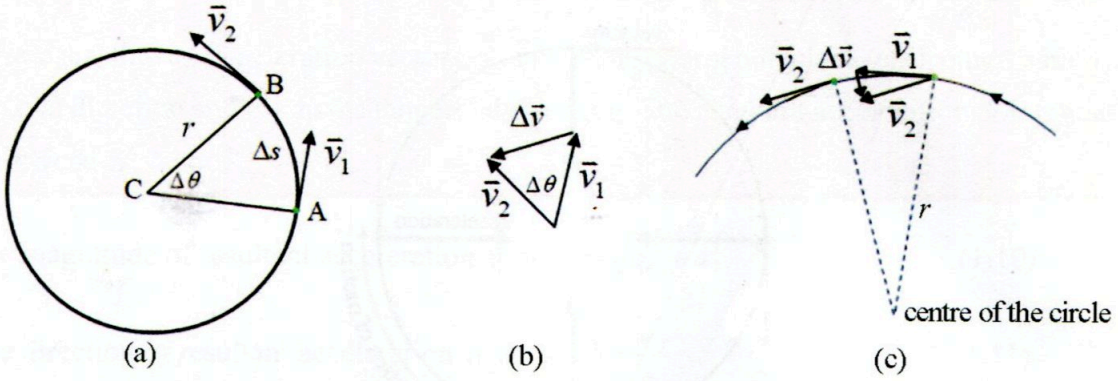


Figure 1.3 (a) Circular motion of an object at constant speed
(b) vector diagram for change in velocity
(c) vector diagram for Δt approaching zero

The magnitude of the centripetal (radial) acceleration can be determined as follow.
The triangle CAB of Figure 1.3(a) is geometrically similar to vector triangle of Figure 1.3(b).

Since $v_1 = v_2 = v$ $$\frac{\Delta v}{v} \approx \frac{\Delta s}{r}$$

When Δt approaches zero, the above equation can be expressed as an equality, so that

$$\Delta v = \frac{v}{r}\Delta s \tag{1.6}$$

The centripetal acceleration a_C can be obtained as

$$a_C = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \lim_{\Delta t \to 0} \frac{v}{r}\frac{\Delta s}{\Delta t} = \frac{v}{r}\lim_{\Delta t \to 0}\frac{\Delta s}{\Delta t}$$

But $\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}$ is the linear speed v of the object moving in circular motion.

Therefore, $$a_C = \frac{v^2}{r} \tag{1.7}$$

Substituting $v = r\omega$ in Eq.(1.7), we get

$$a_C = r\,\omega^2 \tag{1.8}$$

For uniform circular motion, the centripetal acceleration vector points towards the centre of the circular path while the linear velocity vector is tangential to the path.

Hence, the centripetal acceleration and linear velocity are perpendicular to each other at every point in the path as shown in Figure 1.4.

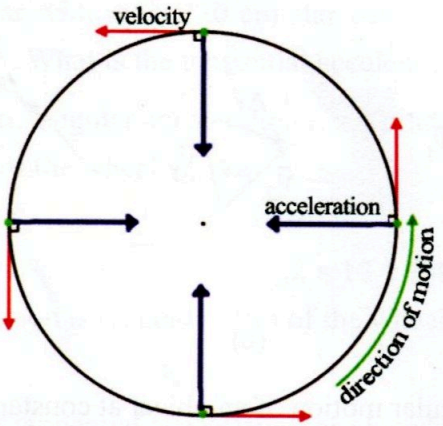


Figure 1.4 Directions of centripetal acceleration and linear velocity for uniform circular motion

Nonuniform Circular Motion

In nonuniform circular motion, an object is moving in a circular path with a varying speed (magnitude of velocity). Therefore, the angular speed of the object is changing and the object experiences the angular acceleration. This angular acceleration gives rise to the tangential acceleration which is tangential to the circle as described previously. In this case, there is tangential acceleration in addition to centripetal acceleration.

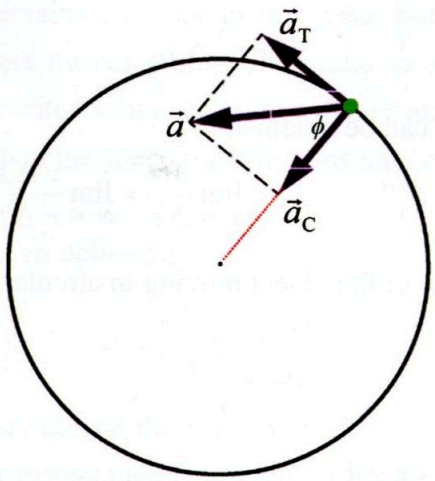


Figure 1.5 Resultant acceleration of nonuniform circular motion

Thus, an object in nonuniform circular motion has a resultant acceleration that is the vector sum of the centripetal and tangential accelerations as shown in Figure 1.5.

$$\vec{a} = \vec{a}_C + \vec{a}_T \qquad (1.9)$$

Unlike tangential acceleration, centripetal acceleration is present in both uniform and nonuniform circular motion.

Note that the two acceleration vectors $\vec{a}_C$ and $\vec{a}_T$ are perpendicular to each other, with $\vec{a}_C$ in the radial direction and $\vec{a}_T$ in the tangential direction. The resultant acceleration points at an angle between $\vec{a}_C$ and $\vec{a}_T$.

The magnitude of resultant acceleration $\vec{a}$ is, $a = \sqrt{a_C^2 + a_T^2}$. (1.10)

The direction of resultant acceleration $\vec{a}$ is, $\tan\phi = \dfrac{a_T}{a_C}$. (1.11)

Example 1.4 The platter of the hard drive of a computer uniformly rotates at 7200 rpm. If the reading head of the drive is located 3 cm from the rotational axis, what is the linear speed and centripetal acceleration of the point on the platter just below the reading head? If a single bit requires 0.5 μm of length along the direction of motion, how many bits per second can the writing head write when it is 3 cm from the axis?

Distance of reading head from the rotation axis $r = 3$ cm $= 0.03$ m,

Angular speed $\omega = 7200 \text{ rpm} = 7200 \times \dfrac{2\pi}{60} = 240\,\pi \text{ rad s}^{-1}$

The linear speed v of the reading head located 3 cm from the rotation axis is,

$$\begin{aligned} v &= r\,\omega \\ &= 0.03 \times 240\,\pi \\ &= 22.62 \text{ m s}^{-1} \end{aligned}$$

The centripetal acceleration a_C is,

$$\begin{aligned} a_C &= \frac{v^2}{r} \\ &= \frac{(22.62)^2}{0.03} = 1.71 \times 10^4 \text{ m s}^{-2} \end{aligned}$$

The distance moved by the writing head in 1 second = 22.62 m

The number of bits written in 1 second $= \dfrac{22.62}{0.5 \times 10^{-6}} = 4.52 \times 10^7$ bits

Example 1.5 A centrifuge has a radius of 20 cm and decelerates from a maximum rotational rate of 7.2×10^3 rpm to rest in 30 s under a constant angular acceleration. It is rotating counter clockwise. What is the magnitude of the resultant acceleration of a point at the tip of the centrifuge at 10 s? What is the direction of the resultant acceleration vector?

Radius of a centrifuge $r = 20$ cm $= 0.2$ m,

Initial angular velocity $\omega_0 = 7.2 \times 10^3 \text{ rpm} = 7.2 \times 10^3 \times \dfrac{2\pi}{60} = 240\,\pi \text{ rad s}^{-1}$,

angular velocity $\omega = 0$, $t = 30$ s

The angular acceleration of centrifuge is,

$$\alpha = \frac{\omega - \omega_0}{t}$$

$$= \frac{0 - 240\,\pi}{30} = -\,8\,\pi \text{ rad s}^{-1}$$

The minus sign means that the centrifuge slows down to be rest.

A point at the tip of centrifuge has both centripetal acceleration and tangential acceleration.

The angular velocity of centrifuge at 10 s,

$$\omega = \omega_0 + \alpha\, t$$

$$= 240\,\pi + (-\,8\,\pi) \times 10 = 160\,\pi \text{ rad s}^{-1}$$

The centripetal acceleration is,

$$a_C = r\,\omega^2$$

$$= 0.2 \times (160\,\pi)^2 = 5.05 \times 10^4 \text{ m s}^{-2}$$

The tangential acceleration is,

$$a_T = r\,\alpha$$

$$= 0.2 \times (-\,8\,\pi)$$

$$= -\,5.03 \text{ m s}^{-2}$$

The magnitude of resultant acceleration is,

$$a = \sqrt{a_C^2 + a_T^2}$$

$$= \sqrt{(5.05\times10^4)^2 + (-5.03)^2} \approx 5.05\times10^4 \text{ m s}^{-2}$$

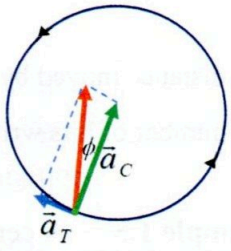
Since $a_C >> a_T$, the magnitude of resultant acceleration is approximately equal to the centripetal acceleration.

The direction of the resultant acceleration vector is,

$$\tan\phi = \frac{a_T}{a_C}$$

$$= \frac{5.03}{5.05\times10^4}$$

$$\phi = 0.0057^\circ$$



The resultant acceleration vector makes angle 0.0057° with $\vec{a}_C$.

Reviewed Exercise

1. Suppose a piece of food is on the edge of a rotating microwave oven plate. Does it experience tangential acceleration, centripetal acceleration, or both when (i) the plate starts to spin faster? (ii) the plate rotates at constant angular velocity? (iii) the plate slows down to stop?

2. A stone tied to a string is moving in a circle of radius 1.5 m at a constant speed 8 m s^{-1}. Calculate the magnitude of the centripetal acceleration of the stone.

Key Words: uniform circular motion, nonuniform circular motion, centripetal acceleration, tangential acceleration

SUMMARY

If the angular velocity of the rotating object changes at a constant rate, then we can say that the angular acceleration is constant and the motion is called the **rotational motion under constant angular acceleration**.

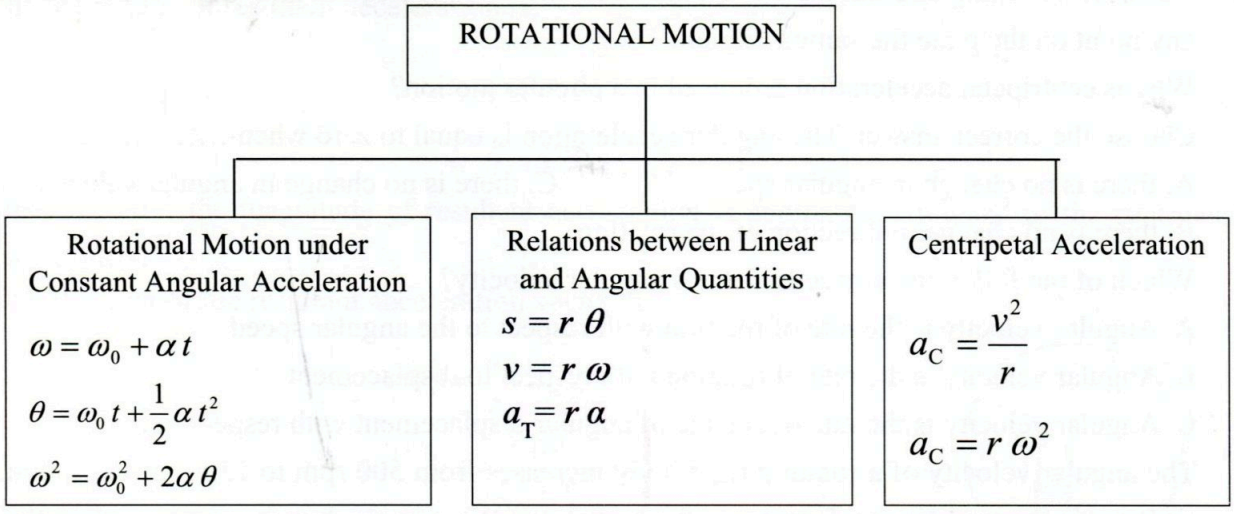
For uniform circular motion, the centripetal acceleration vector points towards the centre of the circular path while the linear velocity vector is tangential to the path. This acceleration is called **centripetal acceleration**.

EXERCISES

1. Consider a rotating circular plate of a lathe machine. Are the angular and tangential speeds at any point on the plate the same? Explain.
2. Why is centripetal acceleration produced in a circular motion?
3. Choose the correct answer. The angular acceleration is equal to zero when .....
   A. there is no change in angular speed
   B. there is no change in direction of the rotation
   C. there is no change in angular velocity
4. Which of the following correctly defines angular velocity?
   A. Angular velocity is the rate of rotation with respect to the angular speed.
   B. Angular velocity is the rate of rotation with respect to displacement.
   C. Angular velocity is the rate of change of angular displacement with respect to time.
5. The angular velocity of a rotating rigid body increases from 500 rpm to 1500 rpm in 2 min. (i) What is the angular acceleration of the body? (ii) What angle does it turn through in this 2 min?
6. A centrifuge rotor is accelerated from rest to 20 000 rpm in 30 s. (i) What is its average angular acceleration? (ii) Through how many revolutions has the centrifuge rotor turned during its acceleration period, assuming constant angular acceleration?
7. A figure skater is spinning with an angular velocity of 15 rad s^{-1}. She then comes to a stop over a brief period of time. During this time, her angular displacement is 5.1 rad. Assuming constant angular acceleration, find (i) her angular acceleration and (ii) the time during which she comes to rest.

8. A tire has a radius of 0.33 m, and its centre moves forward with a linear speed of 15 m s^{-1}. (i) What is angular velocity of the wheel? (ii) Relative to the axle, what is linear speed of a point located 0.175 m from the axle?
9. A child spins a toy top, applying a force to the peg in the middle. The force applied results a tangential acceleration of the peg. If the radius of the peg is 0.5 cm and the tangential acceleration applied is 0.54 m s^{-2}, what is the angular acceleration of the top?
10. A 150 g ball at the end of a string is revolving uniformly in a horizontal circle of radius 0.6 m. The ball makes 2 revolutions in a second. What is its centripetal acceleration?
11. A carousel is initially at rest. It is given a constant angular acceleration 0.06 rad s^{-2}, which increases its angular velocity for 8 s. At t = 8 s, determine the magnitude of the following quantities: (i) the angular velocity of the carousel, (ii) the linear velocity of a child located 2.5 m from the centre, (iii) the tangential acceleration of that child, (iv) the centripetal acceleration of the child, and (v) the total linear acceleration of the child.

CONCEPT MAP

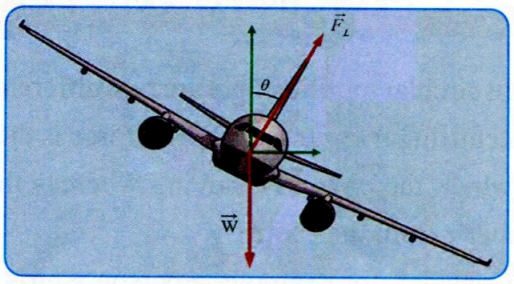


The banked angle θ of the wings of the plane is,

$$\tan\theta = \frac{v^2}{rg}$$

$$= \frac{(222.22)^2}{2\times10^3\times9.8}$$

$$\theta = 68.35^\circ$$



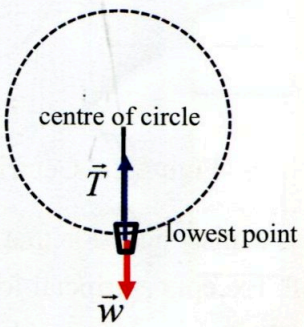
Example 2.2 A 60 cm rope is tied to the handle of a bucket which is then whirled in a vertical circle. The mass of the bucket is 3 kg. If the tension of the rope at the lowest point in its path is 50 N, find the speed of the bucket at that point.

Mass of bucket, $m = 3$ kg

Length of rope = radius of vertical circle, $r = 60$ cm $= 0.6$ m

Force acting on bucket are;

(i) Tension force on bucket $T = 50$ N, and
(ii) Weight of bucket $w = m\,g = 3\times9.8 = 29.4$ N



The net force, $F_{net} = T - w = 50 - 29.4 = 20.6$ N

Since F_{net} provides the centripetal acceleration,

$$F_{net} = F_C = \frac{mv^2}{r}$$

$$20.6 = \frac{3v^2}{0.6}$$

$$v = 2.03 \text{ m s}^{-1}$$

Reviewed Exercise

1. What is the banking angle for an expressway off-ramp curve of radius 50 m at a limiting speed of 50 km h^{-1}?
2. Choose the correct answer.
   A ball attached to a string is whirled around a vertical circle. At the bottom of the arc,
   A. it has no net force.
   B. it has a net force acting upward.
   C. it has a net force acting downward.

Key Words: centripetal acceleration, centripetal force, banking angle

CHAPTER 2

FORCES IN CIRCULAR MOTION

According to Newton's second law, an object that is accelerating must have a net force acting on it. An object moving in a circle, such as a ball on the end of a string, must therefore have a force applied to it to keep it moving in the circle. This force is known as the centripetal force. In this chapter, centripetal force as well as the concepts of centrifugal force will be studied.

Learning Outcomes

It is expected that students will

- examine the centripetal force causing centripetal acceleration.
- investigate the concepts of centrifugal force.
- understand the applications of centripetal and centrifugal forces.

2.1 FORCES CAUSING CENTRIPETAL ACCELERATION

It has been shown in the previous chapter that for uniform circular motion (speed v = constant), the acceleration is centripetal acceleration which is directed towards the centre of the circle at any moment. According to Newton's second law, an object in a circular motion requires a net force directed towards the centre of the circle as shown in Figure 2.1. This net force is called centripetal force. In the absence of this force, the object would not move along a circle but move in a straight line, according to Newton's first law.

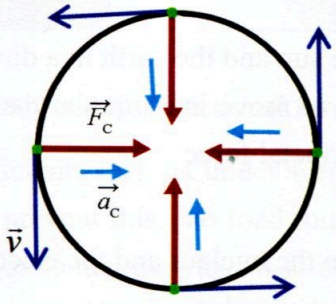


Figure 2.1 Directions of centripetal force, centripetal acceleration and linear velocity in uniform circular motion

Using Newton's second law, $F_{net} = m\,a$, the magnitude of the centripetal force is

$$F_C = m\,a_C = \frac{m\,v^2}{r} \qquad (2.1)$$

Substituting $v = r\omega$ in Eq. (2.1),

$$F_C = m\,r\,\omega^2 \qquad (2.2)$$

For an object to move in a circle, a force is needed to deviate from its straight line path and accelerate towards the centre of the circle. For example, to swing a ball on the end of a string in a circle, we pull on the string and the string exerts the force (tension) on the ball. The tension in the string provides the required centripetal force for the ball to keep moving in circular path.

Some examples of the centripetal force are the gravitational force of the earth on the moon orbiting the earth, the gravitational force of the sun on the planets orbiting the sun, the electric force acting on an electron orbiting the nucleus as illustrated in Figure 2.2.

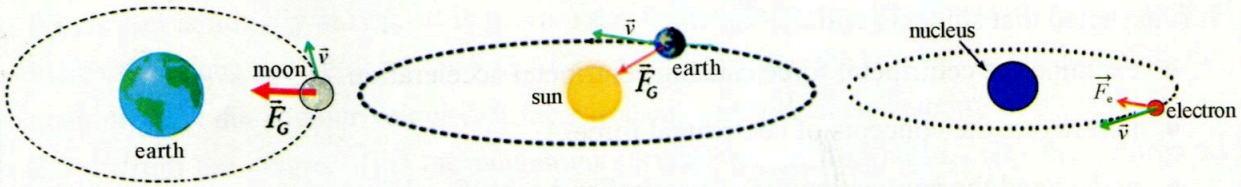


Figure 2.2 Some examples of the centripetal forces

Different Forms of Centripetal Force

There are different forces which can provide centripetal force depending on the motion of bodies under various conditions. A few more examples of the centripetal forces are described as follows.

(i) **Tension force (T)**

When whirling a body attached to a string, a tension force is originated. If this tension force is normal to the direction of motion of a body moving at constant speed, this force causes the body to move in a circular path. Thus, the tension force behaves as a centripetal force.

(ii) **Gravitational force (F_G)**

An attractive force exists between the sun and the earth in a direction perpendicular to the path of the earth. This force causes the earth to move in a circular path around the sun. In this case, the gravitational force provides as a centripetal force.

(iii) **Electric force (F_e)**

The Coulomb attractive force between the nucleus and the electron acts as the required centripetal force for the electron to move around the nucleus.

(iv) **Frictional force (F_f)**

When a car turns in a circular path or a curve, a frictional force exists between the road and the car tires. This frictional force acts normally to the direction of motion of the car, causing the car to move in a curved path. In this case, the frictional force behaves as a centripetal force.

(v) **Normal force (F_N)**

When a car moves in a circular path which is inclined to the horizontal, the reaction force of the road acts normally to the car. Resolving this reaction force (normal force) into two components,

2.2 CENTRIPETAL AND CENTRIFUGAL FORCES

In circular motion, there are two different forces that are encountered. They are centripetal force and centrifugal force. Centripetal force is the force acting towards the centre of the circular path around which the object is moving whereas the centrifugal force appears to act in a direction pointing away from the centre.

Consider, for example, a person whirls a ball on the end of a string in a horizontal circle around his head. The person pulls the string to exert an inward force on the ball. This force is the required centripetal force by the ball to move around the circular path. According to Newton's third law the ball exerts a reaction force on the hand as shown in Figure 2.5.

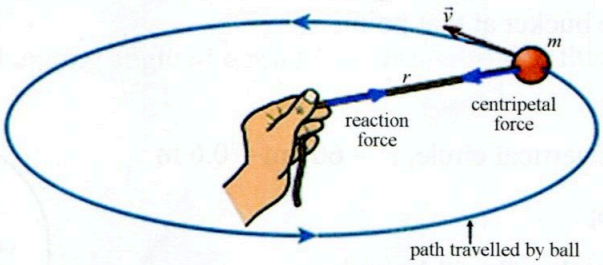


Figure 2.5 Centripetal force on the ball and its reaction force on the hand

There is a misconception that this reaction force is the outward-directed centrifugal force acting on the ball. Except centripetal force, there is no other force acting on the ball along the string.

As another example, consider a person on a moving car. When the car is moving with constant velocity along a straight road, there is no net force on the person. Hence, the person feels no force acting on him. However, when the car moves around a curve the person feels an outward force exerted on him. This is the centrifugal force caused by the effect of inertia.

Actually, the centrifugal force is not a real force. It is an apparent or fictitious force which is equal and opposite to the centripetal force. Although centrifugal force is not a real force, the concept of centrifugal force can be applied in rotating devices, such as centrifuges, washing machines, centrifugal pumps, centrifugal governors and centrifugal clutches.

Centrifuge

A centrifuge is a piece of equipment that puts an object in rotation around a fixed axis (spins it in a circle).

In a laboratory centrifuge that uses sample test tubes, the radial acceleration causes denser particles to settle to the bottom of the tube, while low density substances rise to the top. This effect is known as sedimentation principle.

the horizontal component acts towards the centre of the circle and helps the car to move in a curved path. The horizontal component of normal force is the required centripetal force.

(vi) **Upward thrust (Lifting Force) on a plane (F_L)**

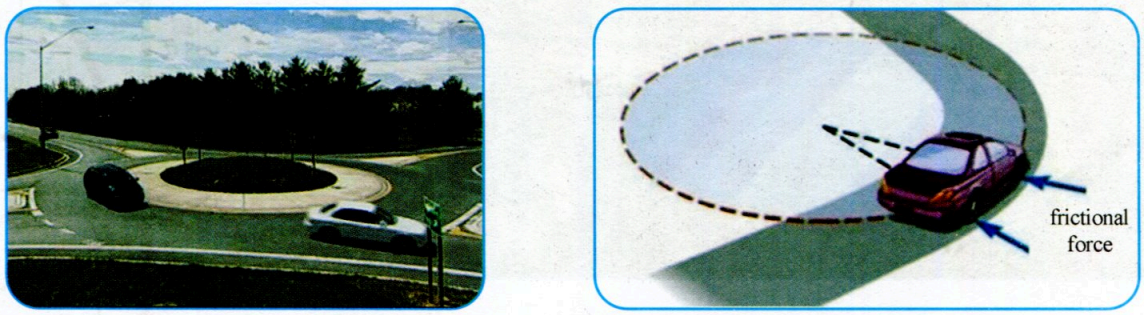
The upward thrust (lifting force) acts normally to an aeroplane body. When the aeroplane tilts, the horizontal component acts normally to the motion of the aeroplane, causing the aeroplane to turn in a curve. This means that the horizontal component of the lifting force on the aeroplane provides the centripetal force.

Banking of Roads

When an automobile drives round a curve, it must have a force exerted towards the centre of the curve. On a flat road (unbanked road), this force is supplied by friction between the tires and the road. As a car makes a turn on unbanked road, the frictional force F_f acting upon the tires of the car provides the required centripetal force for circular motion as shown in Figure 2.3. Hence,

$$F_f = \frac{m v^2}{r} \qquad (2.3)$$

where m is mass of the car, v is speed of the car and r is radius of the path.



(a) (b)

Figure 2.3 (a) Unbanked road (b) Frictional force provides centripetal force

If the frictional force is not sufficient, such as under wet and slippery conditions, the car will skid out of the circular path. To prevent this, the road must be banked to provide the necessary centripetal force to the automobiles so that they take a safe turn without skidding.

Banking of roads is raising the outer edge of the curved road higher than the inner edge. The angle that the inclined track makes with horizontal is called angle of banking. The normal force F_N exerted by a banked road is perpendicular to the road. Horizontal component of this normal force provides the required centripetal force. For every banked curve, there is a limiting speed at which the entire centripetal force is supplied by the horizontal component of the normal force, so that frictional force between the tires and the road is not required.

In this case, the horizontal component of the normal force, $F_N \sin\theta$, towards the centre of the curve of radius r just equals the centripetal force ($\frac{mv^2}{r}$) as shown in Figure 2.4.

$$F_N \sin\theta = \frac{mv^2}{r} \tag{2.4}$$

Since there is no vertical motion, the net force on car in y direction is zero.

$$F_N \cos\theta + (-mg) = 0$$

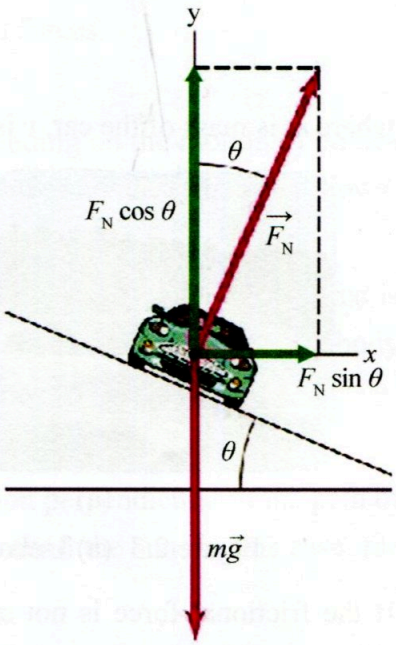
$$F_N \cos\theta = mg$$

Hence,

$$\frac{F_N \sin\theta}{F_N \cos\theta} = \frac{m v^2/r}{m g}$$

$$\tan\theta = \frac{v^2}{rg} \tag{2.5}$$

According to Eq.(2.5), the banking angle of a road θ is chosen so that this condition holds for the limiting speed v.



(a) (b)

Figure 2.4 (a) Banked Road (b) horizontal component of normal force provides centripetal force

Example 2.1 A jet plane is flying around the airport with a speed of 800 km h^{-1} along a circular path with a radius of 2 km as shown in given figure. At what angle must the wings of the plane be banked?

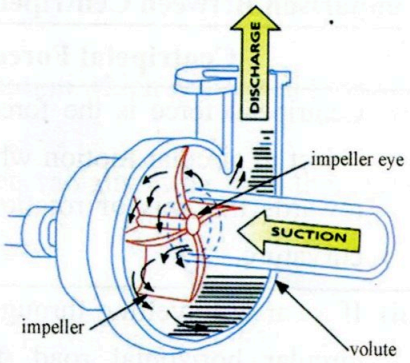
Speed of jet plane v = 800 km h^{-1} = 222.22 m s^{-1}

Radius of circular path $r = 2 \text{ km} = 2 \times 10^3$ m

The wings of the plane must be banked an angle θ with vertical.

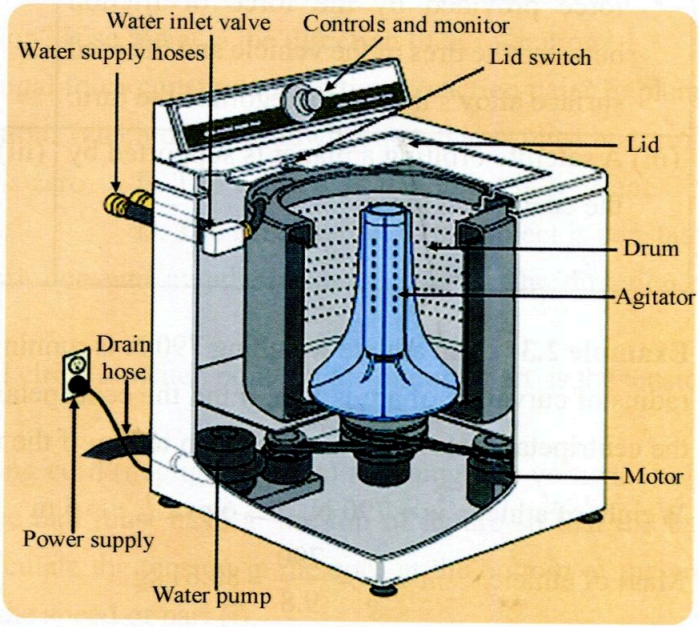
Centrifugal Pump

Centrifugal pump is a hydraulic machine which converts mechanical energy into hydraulic energy (i.e. pressure energy) by the use of centrifugal force acting on the fluid. The rotating impeller moves water or other fluids by using centrifugal force.



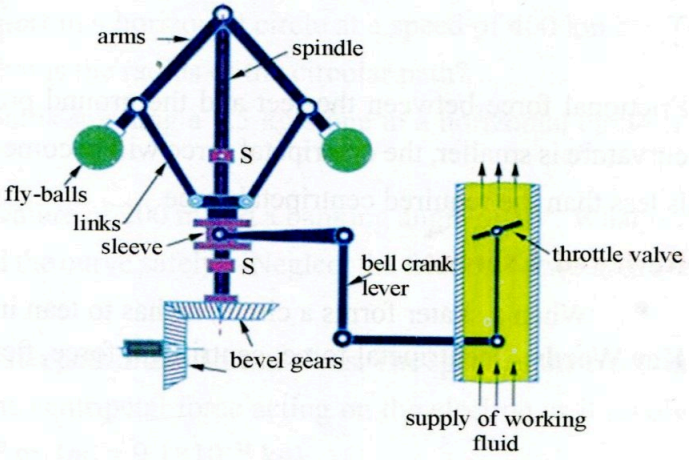
Washing Machine

The centrifugal action of a washing machine is used in spin drying the clothes. When all the clothes are completely washed, they are spun very fast and excess water is removed completely with the help of centrifugal force.



Centrifugal Governor

The centrifugal governor is a device used to measure and regulate the speed of a machine, such as an engine. The centrifugal governors consist of two balls of equal mass, which are attached to the arms. As the engine is working, these balls also revolve around the spindle at a certain speed and maintain at a position due to the centrifugal force.



Comparison Between Centripetal and Centrifugal Force

Centripetal Force	Centrifugal Force
(i) Centripetal force is the force acting on an object in circular motion which is directed towards the axis of rotation or centre of curvature.	(i) Centrifugal force in a circular motion is not a real force; it acts radially away from the centre of the circle. It results from the effect of inertia.
(ii) If a car is travelling through a curve on a circular horizontal road, the centripetal force provided by the force of friction between the tires of the vehicle and the road surface allows the car to negotiate the turn.	(ii). When a car in motion takes a sudden turn in a curved path, passengers in the car experience an outward push. This is due to the centrifugal force acting on passengers.
(iii) A satellite orbiting a planet is supported by the centripetal force.	(iii) Removing the excess water by spin drying clothes in a washing machine is with the help of centrifugal force.

Example 2.3 An athlete weighing 790 N is running a curve at a speed of 6 m s^{-1} in an arc. The radius of curvature of arc is 5 m. Find the centripetal force acting on him. Which force provides the centripetal force? What will happen to him if the radius of curvature is smaller?

Weight of athlete, $w = 790$ N, $v = 6$ m s^{-1}, $r = 5$ m

Mass of athlete, $m = \frac{w}{g} = \frac{790}{9.8} = 80.61$ kg

The required centripetal force to run a curve, $F_C = \frac{mv^2}{r}$

$$= \frac{80.61(6)^2}{5} = 580.4 \text{ N}$$

Frictional force between the feet and the ground provides the centripetal force. If the radius of curvature is smaller, the centripetal force will become greater. He will slip since the frictional force is less than the required centripetal force.

Reviewed Exercise

- When a skater forms a circle, he has to lean inwards. Explain with a force diagram.

Key Words: centripetal force, centrifugal force, fictitious force

SUMMARY

Centripetal force is the force acting on an object in circular motion which is directed towards the axis of rotation (or) centre of curvature.

Centrifugal force in a circular motion is not a real force; it acts radially away from the centre of the circle. It results from the effect of inertia.

EXERCISES

1. Imagine a car driving over a hillock at a constant speed. Once the car has reached the apex of the hillock, does it have the acceleration? If so, what is the direction of acceleration?
2. Does the centripetal force and centrifugal force constitute an action - reaction pair? Explain.
3. An object is in a uniform circular motion. Which of the following statements must be true?
   A. The net force acting on the object is zero. B. The velocity of the object is constant.
   C. The speed of the object is constant. D. The acceleration of the object is constant.
4. For an object moving in a circle with constant angular velocity, what is the direction of acceleration of the object?
5. When a bob is swinging in the vertical circle at which point on the circular part is the tension in the rope the greatest?
6. A 0.15 kg ball on the end of 1.1 m long cord (negligible mass) is swung in a vertical circle. (i) Determine the minimum speed the ball must have at the top of its arc so that the ball continues moving in a circle. (ii) Calculate the tension in the cord at the bottom of the arc, assuming the ball is moving at twice the speed of part (i).
7. An 8.3 kg mass is attached to a string that has a breaking strength of 1500 N. If the mass is whirled in a horizontal circle of radius 80 cm, what maximum speed can it have?
8. An aeroplane is circling above an airport in a horizontal circle at a speed of 400 km h^{-1}. The banking angle of the wings is 20°. What is the radius of the circular path?
9. What is the tension in a 1 m string that is spinning a 0.5 kg stone in a horizontal circle with 3 rps?
10. A curved roadway has a radius of curvature of 200 m and a banking angle of 10°. What is the highest speed at which a car can round the curve safely? (Neglect the friction between tire and the road)
11. In an atom, an electron moves in a circular path around the nucleus. The speed of the electron is approximately 2.2×10^{6} m s^{-1}. Find the centripetal force acting on the electron as it revolves in a circular orbit of radius 0.53×10^{-10} m. ($m_e = 9.1\times10^{-31}$ kg)

Using continuity equation, $A_{pipe}\, v_i = A_{nozzle}\, v_{nozzle}$

$$4.91 \times 101.83 = 0.5 \; v_{nozzle}$$

$$v_{nozzle} = 1000 \text{ cm s}^{-1}$$

Example 3.2 What area must a heating duct have if air moving 3 m s^{-1} along it can replenish the air every 15 min in a room of volume 300 m^3? Assume the density of air remains constant.

Let the area of heating duct be A.

The velocity of air at duct $v = 3$ m s^{-1}

The volume of air coming in the room in 15 min = volume of room

$$= 300 \text{ m}^3$$

$$\text{Volume flow rate } \frac{V}{t} = A\,v$$

$$\frac{300}{15 \times 60} = A \times 3$$

$$A = 0.11 \text{ m}^2$$

Example 3.3 Water runs into a fountain, filling all the pipes, at a steady rate of 0.75 m^3 s^{-1}. (i) How fast will it shoot out of a hole 4.5 cm in diameter? (ii) At what speed will it shoot out if the diameter of the hole is three times as large?

Volume flow rate $\frac{V}{t} = 0.75\,\text{m}^3\text{s}^{-1}$, radius of the hole $= \frac{4.5}{2} = 2.25\,\text{cm} = 0.0225\,\text{m}$

(i) $$\frac{V}{t} = A\,v = \pi r^2 v$$

$$0.75 = 3.142 \times (0.0225)^2 v$$

$$v = 471.51 \text{ m s}^{-1}$$

(ii) Since diameter is three times larger, area is nine times larger. $A_2 = 9\,A_1$

Using continuity equation,

$$A_1 v_1 = A_2 v_2$$

$$v_2 = \frac{v_1}{9}$$

Therefore speed of water shooting out $v_2 = \frac{v_1}{9} = \frac{471.51}{9} = 52.39 \text{ m s}^{-1}$

Example 3.4 Water tank of dimension 3 m × 3 m × 3 m and the base of the tank is 10 m above the ground. It takes 1 h to fill. (i) Find the power output of the pump to fill the tank. (ii) Find the flow rate through the pipe and the speed of the water flow in the pipe which has 5 cm radius.

CONCEPT MAP

FORCES IN CIRCULAR MOTION

Centripetal Force

- centripetal force is the force acting towards the centre of the circular path around which the object is moving.

$$F_C = m\,a_C = \frac{m\,v^2}{r}$$

Centrifugal Force

- centrifugal force appears to act in a direction pointing away from the centre.

Applications of centripetal forces

- road banking
- the gravitational force on the planets
- electrical force on an electron moving around the nucleus

Applications of centrifugal forces

- rotating devices; centrifuges, washing machines, centrifugal pumps, centrifugal governors, etc.

Comparison between centripetal and centrifugal forces

- a satellite orbiting a planet is supported by the centripetal force.
- removing the excess water by spin drying clothes in a washing machine is with the help of centrifugal force.

CHAPTER 3

FLUID DYNAMICS

Fluid is a substance that offers no permanent resistance to deforming forces. The study of fluids has two parts, statics and dynamics. Fluid statics concentrates on fluids at rest, while fluid dynamics concerns with the fluid in motion.

We have already studied fluid statics in Grade 10 and Grade 11; which includes atmospheric pressure, liquid pressure, Archimedes' principle and Pascal's law. Now we are going to study fluid dynamics or fluids in motion. Motion of fluids is an important application in our daily life; such as river flow, water distribution systems, gas pipelines, aviation (travels in air) and motion underneath the surface of water.

The study of the flow of liquid is called hydrodynamics, whereas hydrostatic is the study of liquid at rest. This chapter mainly concerns with hydrodynamics; however a section on surface tension and capillarity, which are important concepts of hydrostatic, is included at the end of the chapter.

Learning Outcomes

It is expected that students will

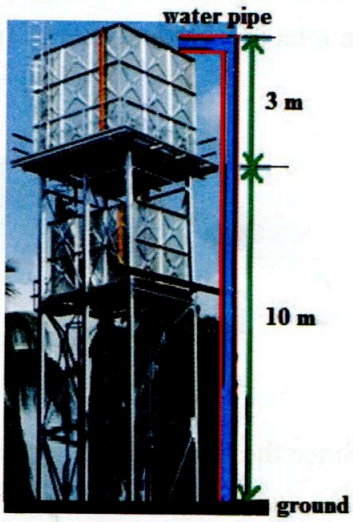
- differentiate fluid flow as laminar and turbulent.
- study equation of continuity for a fluid motion.
- understand Bernoulli's principle as a relationship between pressure energy, kinetic energy and potential energy of a fluid in streamline (laminar) flow.
- realize some manifestations of Bernoulli's principle.
- understand viscosity and Stokes' law.
- distinguish between cohesion and adhesion and explain surface tension in liquids.
- discuss capillarity.

3.1 LAMINAR AND TURBULENT FLOW

Fluid is either liquid or gas. The volume of a gas depends on pressure and temperature obeying Boyle's law and Charles' law. However, the volume of a liquid does not depend on pressure if the temperature is constant. Hence, the liquid is incompressible (i.e. density of liquid is constant for any pressure provided that the temperature is kept constant).

In the flow of fluids, the path or flow of the fluid particles is called streamline as shown in Figure 3.1 (a). Motion of fluids can be classified by two types, laminar flow and turbulent flow.

$V = 3 \times 3 \times 3 = 27\ m^3$, $h = 10 + 3 = 13$ m, $t = 1$ h $= 3600$ s



(i) Power, $P = \frac{W}{t} = \frac{m\,g\,h}{t} = \frac{\rho\,V\,g\,h}{t}$

$$= \frac{1000 \times 27 \times 9.8 \times 13}{3600} = 955.5\ W$$

(ii) $r = 5$ cm $= 5 \times 10^{-2}$ m

Flow rate $\frac{V}{t} = \frac{27}{3600} = 7.5 \times 10^{-3} m^3 s^{-1}$

Flow rate $\frac{V}{t} = A\,v$

$7.5 \times 10^{-3} = \pi\,(5 \times 10^{-2})^2\,v$

$v = 0.96\ m\ s^{-1}$

Reviewed Exercise

1. Why does the river flow faster in passing the defile (narrow river passage)?
2. A shower head has 20 circular openings, each with radius 1 mm. The shower head is connected to a pipe with radius 0.8 cm. If the speed of water in the pipe is 3 m s^{-1}, what is its speed as it exits the shower-head openings?

Key Words: flow rate, cross-sectional area, flow speed, continuity equation

3.3 BERNOULLI'S EQUATION

Bernoulli's equation, in fluid dynamics, relates the pressure, flow speed, and height for the steady (laminar) flow of an ideal incompressible fluid. The internal friction of the fluid flow is negligible (i.e. no viscosity). This equation was first derived in 1738 by the Swiss mathematician Daniel Bernoulli.

Bernoulli's equation can be derived using the work-energy principle to a fluid flowing in a tube. Consider a section of the fluid flowing in the tube with uneven cross section as shown in Figure 3.3. This fluid section is initially between two cross sections a and c. After a small time interval the fluid at a moves to b, and the fluid at c moves to d. The cross-sectional area, speed and pressure at two ends of the fluid section are A_1, v_1, p_1 and A_2, v_2, p_2, respectively.

Daniel Bernoulli
(1700-1782)

1. Laminar Flow

If fluid particles are moving steady in smooth paths in layers, with each layer moving smoothly past the adjacent layers with no mixing; such a flow is called laminar flow. In this steady laminar flow, streamlines do not cross each other and every fluid particle arriving at a given point has the same velocity as illustrated in Figure 3.1 (b).

2. Turbulent Flow

If the flow or path of the fluid particles are irregular, their direction is always changing or whirling; this fluid movement is called turbulent flow. In turbulent flow the speed of the fluid at a point is continuously undergoing changes in both magnitude and direction as shown in Figure 3.1 (c).

Ideal laminar fluid flow considers the following facts.

- The fluid is incompressible.
- The flow has no friction or friction is neglected. (Internal friction in fluids is called viscosity; and hence, zero viscosity.)
- The fluid flow is laminar (steady) and turbulent flow is not taken into account. (For example, sea breeze is a laminar flow and storm is a turbulent flow.)

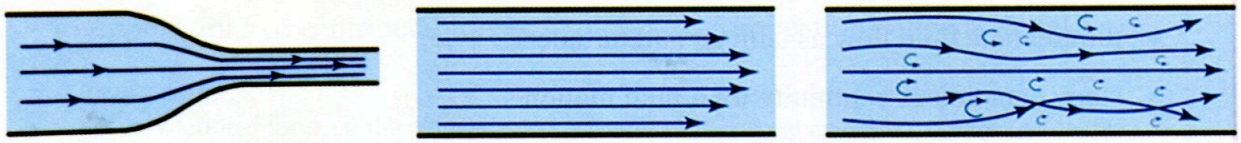


Figure 3.1 (a) Stream lines (b) laminar flow (c) turbulent flow

Reviewed Exercise

1. Is motor car brake oil compressible or non-compressible?
2. Are gases compressible or non- compressible?
3. What is the difference between laminar flow and turbulent flow?
4. At a river junction, water streams are whirling; is it laminar or turbulent?

Key Words: fluid, incompressible, laminar, turbulent, streamline, viscosity

3.2 EQUATION OF CONTINUITY FOR FLUIDS

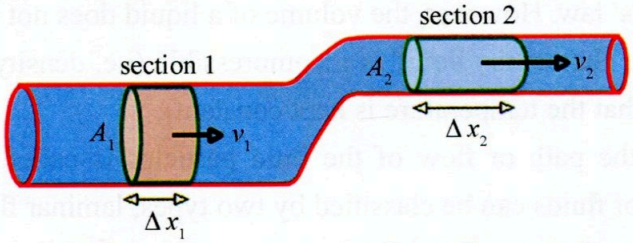


Figure 3.2 Illustrating constant volume flow rate in a pipe

Consider an ideal fluid there must be laminar flow and no viscosity. The density of fluid is constant. In Figure 3.2 all the liquid that flow through section 1 (larger cross sectional area A_1) flow through section 2 (smaller cross sectional area A_2). The mass flowing in section 1 must equal the mass flowing in section 2. Hence, volume flow rate $\frac{V}{t}$ is the same throughout the pipe.

$$m_1 = m_2$$

$$\rho V_1 = \rho V_2$$

Therefore,
$$\frac{V_1}{t} = \frac{V_2}{t}$$

Since $V_1 = A_1 \Delta x_1$ and $V_2 = A_2 \Delta x_2$

$$\frac{A_1 \Delta x_1}{t} = \frac{A_2 \Delta x_2}{t}$$

$$A_1 v_1 = A_2 v_2 \qquad (3.1)$$

where v_1 and v_2 are speed of liquid at section 1 and section 2 respectively.
The equation (3.1) states that the liquid entering one end of the pipe leaves from the other end with a constant volume flow rate.
This equation is known as equation of continuity for fluid flow.

The larger the area of the pipe, the slower is the speed; and the smaller the area, the faster is the speed. For example, the speed of water spraying from a hose increases by making the hose area smaller.

Example 3.1 A gardener uses a water pipe of 2.5 cm diameter. It takes 1 min to fill 30 litre bucket. What is the initial speed of water coming out of the hose? Then nozzle with an opening 0.5 cm^2 is attached to the hose. Find the water speed coming out from the nozzle.

Initial pipe area, $A_{\text{pipe}} = \pi r^2 = \pi \left(\frac{2.5}{2}\right)^2 = 4.91 \text{ cm}^2$

Volume of water flow in 1 min (60 s) V= 30 litre = 30 × 1000 cm^3

Volume flow rate $\frac{V}{t} = \frac{30 \times 1000}{60} = 500 \text{ cm}^3 \text{ s}^{-1}$ (this value is constant for the system)

Since volume flow rate $\frac{V}{t} = A v$,

The initial speed of water coming out of the hose $v_i = \frac{V/t}{A_{\text{pipe}}}$

$$= \frac{500}{4.91} = 101.83 \text{ cm s}^{-1}$$

When nozzle is attached, nozzle area $A_{\text{nozzle}} = 0.5 \text{ cm}^2$

According to work-energy principle, the work done on this section of fluid must equal the change in total mechanical energy (i.e. kinetic energy + potential energy).

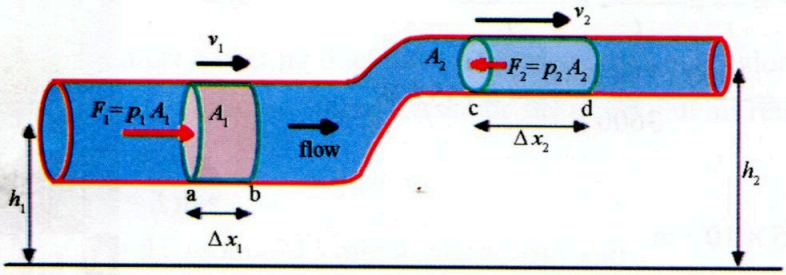


Figure 3.3 A section of a tube of flowing fluid

Since the fluid is incompressible, $A_1 \Delta x_1 = A_2 \Delta x_2 = \Delta V$

The work done on the fluid in this case is due to the pressure of the fluid.

The work done $\Delta W = F_1 \Delta x_1 - F_2 \Delta x_2$

$$= p_1 A_1 \Delta x_1 - p_2 A_2 \Delta x_2$$

$$= p_1 \Delta V - p_2 \Delta V = (p_1 - p_2) \Delta V$$

(The second term has a negative sign because the force at c opposes the force at a.)

The mass of the fluid between a and b $= \rho A_1 \Delta x_1 = \rho \Delta V$

The mass of the fluid between c and d $= \rho A_2 \Delta x_2 = \rho \Delta V$

The change in kinetic energy $\Delta KE = \frac{1}{2} \rho \Delta V (v_2^2 - v_1^2)$

The change in potential energy $\Delta PE = \rho \Delta V g (h_2 - h_1)$

Using the work-energy principle $\Delta W = \Delta KE + \Delta PE$

$$(p_1 - p_2) \Delta V = \frac{1}{2} \rho \Delta V (v_2^2 - v_1^2) + \rho \Delta V g (h_2 - h_1)$$

$$(p_1 - p_2) = \frac{1}{2} \rho (v_2^2 - v_1^2) + \rho g (h_2 - h_1)$$

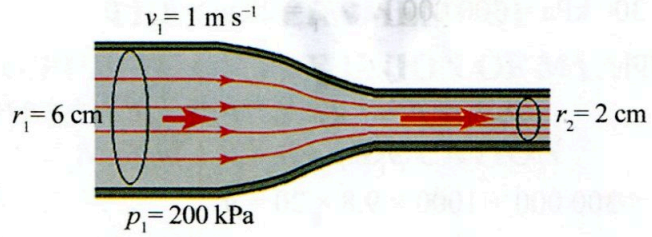
$$p_1 + \frac{1}{2} \rho v_1^2 + \rho g h_1 = p_2 + \frac{1}{2} \rho v_2^2 + \rho g h_2$$

Thus, $$p + \frac{1}{2} \rho v^2 + \rho g h = \text{constant} \qquad (3.2)$$

This equation is known as Bernoulli's equation.

In words, Bernoulli's theorem states that the sum of pressure energy, kinetic energy and potential energy per unit volume of an incompressible, non-viscous fluid in a streamlined laminar flow remains constant.

Example 3.5 Water flowing through a restriction in a horizontal pipe is shown in the given figure. The radius of the pipe at the left end and right end are 6 cm and 2 cm respectively. If the velocity and pressure of water at the left end of pipe are 1 m s^{-1} and 200 kPa, find the velocity and pressure of water at the right end of the pipe.



Let velocity and pressure at right end be v_2, p_2.

Using continuity equation,

$$A_1 v_1 = A_2 v_2$$

$$\pi r_1^2 v_1 = \pi r_2^2 v_2$$

$$v_2 = v_1 (\frac{r_1}{r_2})^2$$

$$= 1 \times \left(\frac{0.06}{0.02} \right)^2 = 9 \text{ m s}^{-1}$$

Using Bernoulli's theorem,

$$p_1 + \frac{1}{2} \rho \ v_1^2 + \rho \ g \ h_1 = p_2 + \frac{1}{2} \rho \ v_2^2 + \rho \ g \ h_2$$

Since $h_1 = h_2$,

$$p_1 + \frac{1}{2} \rho \ v_1^2 = p_2 + \frac{1}{2} \rho \ v_2^2$$

$$p_2 = p_1 - \frac{1}{2} \rho \ (v_2^2 - v_1^2)$$

$$= 200 \times 10^3 - \frac{1}{2} \times 1000 \ (9^2 - 1^2)$$

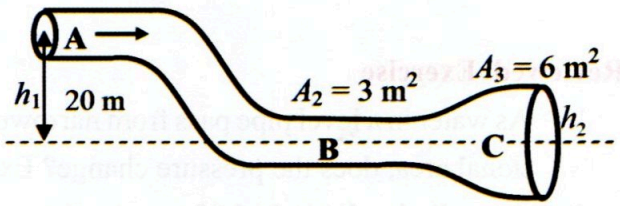
$$= 160 \times 10^3 \text{ Pa} = 160 \text{ kPa}$$

Example 3.6 Calculate the pressure and speed of water at points B and C shown below. (density of water is 1000 kg m^{-3})

$A_1 = 3$ m^2

$v_1 = 10$ m s^{-1}

$p_1 = 300$ kPa



Using continuity equation,

$$A_1 v_1 = A_2 v_2$$

Since $A_1 = A_2$, $v_1 = v_2 = 10$ m s^{-1}

Using continuity equation

$$A_2 v_2 = A_3 v_3$$

$$3 \times 10 = 6 \ v_3$$

$$v_3 = 5 \text{ m s}^{-1}$$

For points A and B, $p_1 = 300$ kPa $= 300\,000$ Pa, $h_1 = 20$ m, $h_2 = 0$

Using Bernoulli's theorem,

$$p_1 + \tfrac{1}{2}\rho v_1^2 + \rho g h_1 = p_2 + \tfrac{1}{2}\rho v_2^2 + \rho g h_2$$

Since $v_1 = v_2$,

$$p_1 + \rho g h_1 = p_2 + \rho g h_2$$

$$300\,000 + 1000 \times 9.8 \times 20 = p_2$$

$$p_2 = 496\,000 \text{ Pa} = 496 \text{ kPa}$$

For points B and C, $v_2 = 10$ m s^{-1} and $v_3 = 5$ m s^{-1}, $p_2 = 496\,000$ Pa

Using Bernoulli's theorem,

$$p_2 + \tfrac{1}{2}\rho v_2^2 + \rho g h_2 = p_3 + \tfrac{1}{2}\rho v_3^2 + \rho g h_3$$

Since $h_2 = h_3$,

$$p_2 + \tfrac{1}{2}\rho v_2^2 = p_3 + \tfrac{1}{2}\rho v_3^2$$

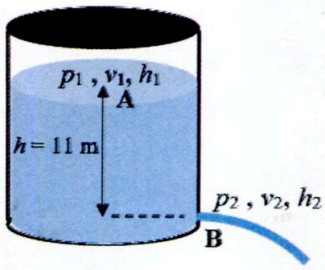
$$496\,000 + \tfrac{1}{2} \times 1000 \times 10^2 = p_3 + \tfrac{1}{2} \times 1000 \times 5^2$$

$$p_3 = 533\,500 \text{ Pa} = 533.5 \text{ kPa}$$

Example 3.7 A sealed tank containing seawater to a height of 11 m also contains air above the water at a pressure of 4 atm. Water flows out from the bottom through a small hole. How fast is this water moving?

At point A (top water surface)

$$p_1 = 4 \text{ atm} = 4 \times 1.01 \times 10^5 \text{ Pa} = 4.04 \times 10^5 \text{ Pa}$$



Since cross-sectional area of water tank is much greater than area of small hole, water level in the tank drops very slowly, we take $v_1 \approx 0$, $h_1 = 11$ m.

At point B (small hole at the bottom)

$p_2 = 1$ atm $= 1.01 \times 10^5$ Pa, $h_2 = 0$ m

For points A and B, using Bernoulli's theorem,

$$p_1 + \tfrac{1}{2}\rho v_1^2 + \rho g h_1 = p_2 + \tfrac{1}{2}\rho v_2^2 + \rho g h_2$$

$$4.04 \times 10^5 + 1000 \times 9.8 \times 11 = 1.01 \times 10^5 + \tfrac{1}{2} \times 1000\, v_2^2$$

$$4.11 \times 10^5 = 500\, v_2^2$$

$$v_2 = 28.66 \text{ m s}^{-1}$$

Reviewed Exercise

1. As water in a level pipe pass from narrower cross-sectional area of pipe to a wider cross-sectional area, does the pressure change? Explain.
2. A small circular hole 6.00 mm in diameter is cut in the side of a large water tank, 14.0 m below the water level in the tank. The top of the tank is open to the air. Find (i) the speed of efflux of the water, and (ii) the volume discharged per second.

Key Words: work-energy principle, incompressible, non-viscous fluid

3.4 MANIFESTATIONS OF BERNOULLI'S THEOREM

1. Lift on the Wings of an Aeroplane

The air flow moving above the top curved surface of the wing has a longer distance to travel and needs to go faster to have the same transit time as the air travelling along the lower flat surface (horizontal position). Hence, the velocity of air flow is high along the top surface and low along the bottom surface. This causes a low pressure above the wing and a high pressure below the wing which provides a lift on the wing.

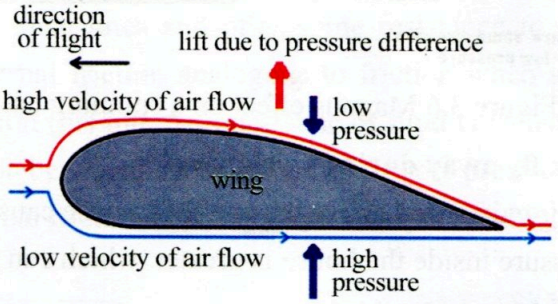


Figure 3.4 Lift on the wings of an aeroplane

2. Insect Sprayer

When the piston of the sprayer is pressed, the fast moving air causes a low pressure at the tip of the nozzle. The liquid inside the container is at a higher pressure, and therefore rushes up through the stem and out the nozzle.

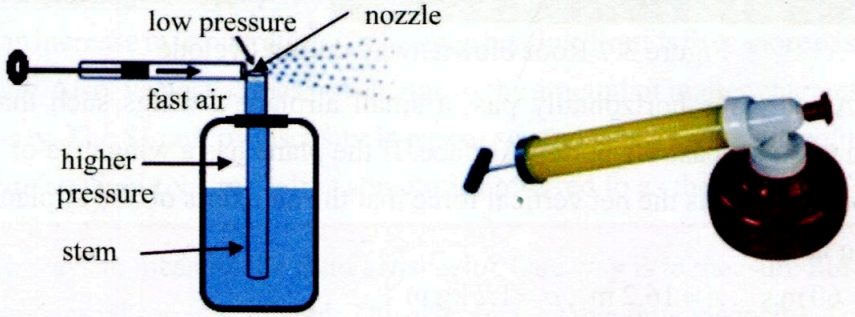


Figure 3.5 The insect sprayer

3. Magnus Effect on a Spinning Ball

The Magnus effect is a particular manifestation of Bernoulli's theorem that is commonly associated with a spinning object moving through a fluid. In the case of a ball spinning through the air, the turning ball drags some of the air around with it. This causes the difference in pressure of the air on opposite sides of the spinning ball and produces the deflection of the ball. The effect is named after the German physicist Heinrich Gustav Magnus, who described the effect in 1852.

3.6 SURFACE TENSION AND CAPILLARITY

1. Surface Tension

Surface tension is the tendency of the surface of a liquid to behave like a stretched elastic skin (or membrane). It is due to the attractive forces between liquid molecules called cohesion. Cohesion refers to forces between molecules of the same substance. For example, cohesion makes water molecules pull together making the water surface to stretch like an elastic skin.

Surface tension is the reason a water strider and it is able to stride along water surface without sinking. Paperclips, razor blades and light coins are also able to rest on water without sinking although their density is higher than that of water as shown in Figure 3.9.

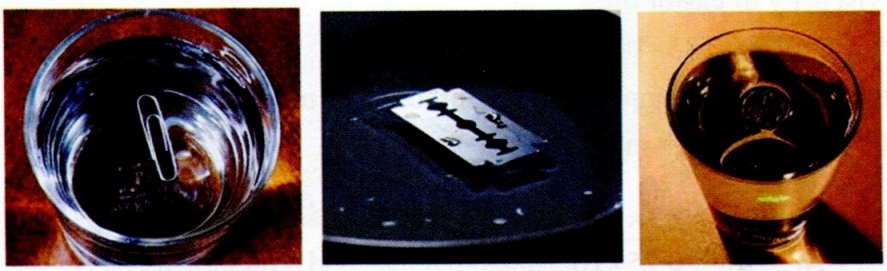


Figure 3.9 Effect of surface tension

Surface tension is defined as the force per unit length that acts across any line on a surface tending to pull the surface apart.

$$\gamma = \frac{F}{l} \tag{3.6}$$

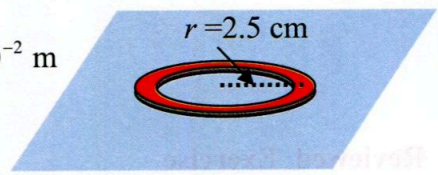
where F = force, l = length, γ = surface tension

The SI unit of surface tension is newton per metre (N m^{-1}).

> Surface tension can also be referred to as a force because it is the tension between liquid molecules.

Example 3.11 A horizontal circular loop of wire has a diameter of 5 cm and is lowered in a sample of crude oil. The additional force required to pull the loop out of the oil is 0.04 N; calculate the surface tension of the crude oil.

$F = 0.04$ N, $d = 5$ cm $= 5\times10^{-2}$ m, $r = \frac{d}{2} = \frac{5}{2} = 2.5$ cm $= 2.5\times10^{-2}$ m



Circumference of the loop $l = 2\,\pi r$

$= 2 \times 3.142 \times 2.5 \times 10^{-2} = 0.16$ m

Additional force F required to pull the loop is the force due to surface tension. Since the circular loop has two rims (outer and inner rims) in contact with the liquid surface,

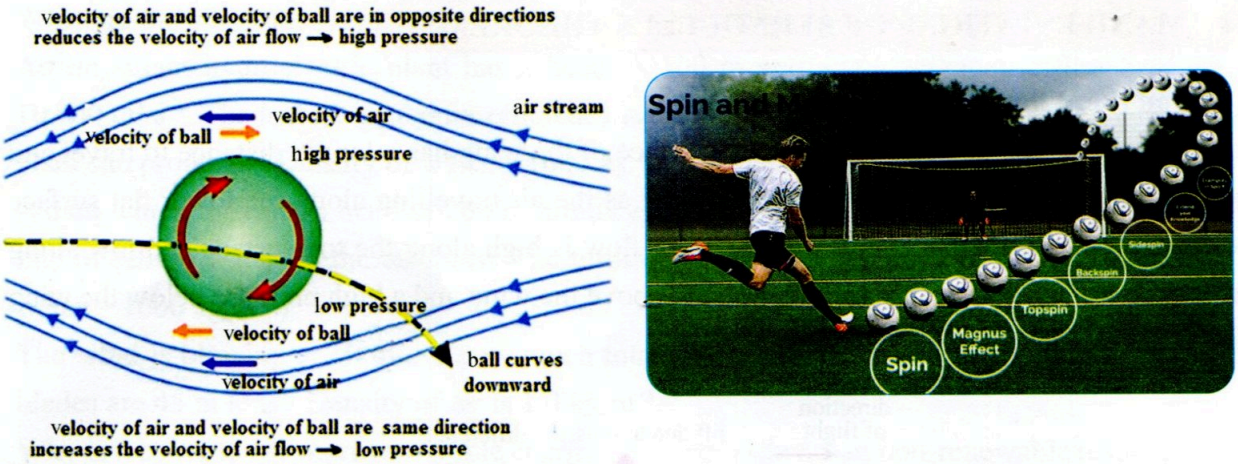


Figure 3.6 Magnus effect on a spinning ball

4. The Roof of Houses can fly away during Cyclones

During cyclones, moving air gain speed above the roof of a house causing the reduced pressure on the roof. Therefore the pressure inside the house is greater, which can raise the roof.

Figure 3.7 Roof blown away during cyclone

Example 3.8 Air streams horizontally past a small airplane's wings such that the speed is 70 m s^{-1} over and 60 m s^{-1} past the bottom surface. If the plane has a wing area of 16.2 m^2 on the top and on the bottom, what is the net vertical force that the air exerts on the airplane? The density of the air is 1.2 kg m^{-3}.

$v_1 = 70\ \text{m s}^{-1}, v_2 = 60\ \text{m s}^{-1}, A = 16.2\ \text{m}^2, \rho = 1.2\ \text{kg m}^{-3}$

For top and bottom surfaces of the wing, using Bernoulli's theorem,

$$p_1 + \frac{1}{2}\rho\ v_1^2 + \rho\ g\ h_1 = p_2 + \frac{1}{2}\rho\ v_2^2 + \rho\ g\ h_2$$

Pressure difference, $p_2 - p_1 = \frac{1}{2}\rho v_1^2 - \frac{1}{2}\rho v_2^2 \qquad (h_1 = h_2)$

$$= \frac{1}{2} \times 1.2 \left[(70)^2 - (60)^2\right] = 780\ \text{Pa}$$

$$F = p\,A$$

$$= 780 \times 16.2 = 12\,636\ \text{N}$$

Reviewed Exercise

1. Can fluid flow from lower pressure to higher pressure?
2. Air is streaming past a horizontal aeroplane wing such that its speed is 120 m s^{-1} at the upper surface and 90 m s^{-1} at the lower surface. If the wing is 10 m long and 2 m wide and density of air is 1.3 kg m^{-3}, find the net lift on it.

Key Words: ideal fluid, kinetic energy per unit volume, potential energy per unit volume

3.5 VISCOSITY (Fluid Friction)

Most of the fluids are not ideal ones and offer some resistance to motion. This resistance to fluid motion is like an internal friction analogous to friction when a solid move on a surface. The resistance to fluid motion (or) internal resistance of fluid is called viscosity. Viscosity arises when there is relative motion between layers of the fluid. More precisely, it measures resistance to flow arising due to the internal friction between the fluid layers as they slip past one another when fluid flows.

Viscosity can also be thought of as a measure of a fluid's thickness (stickiness) or its resistance to objects passing through it. A fluid with large viscosity resists motion because its strong intermolecular forces give it a lot of internal friction, resisting the movement of layers past one another. On the contrary, a fluid with low viscosity flows easily because its molecular make up results in very little friction when it is in motion. Gases also exhibit viscosity, but it is harder to notice in ordinary circumstances.

The viscosity of liquids decreases rapidly with an increase in temperature, and the viscosity of gases increases with an increase in temperature. Thus, upon heating, liquids flow more easily, whereas gases flow more slowly. Also, viscosity does not change as the amount of matter changes, therefore it is an intensive property. The SI unit of viscosity is pascal second (Pa s) which is equivalent to newton second per metre squared (N s m^{-2}). It is sometimes referred to as the poiseuille (Pl).

There are two ways to measure the fluid's viscosity. One way is to measure fluid's resistance to flow when an external force is applied. This is known as dynamic viscosity. And the other way is to measure the resistive flow of a fluid under the weight of gravity. It is known as kinematic viscosity, which is more useful than dynamic viscosity (absolute viscosity).

Newton's Law of Viscosity

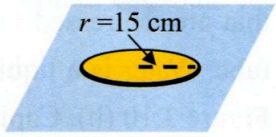
Newton's law of viscosity states that the shear stress between the two adjacent layers of fluid is directly proportional to the negative value of the velocity gradient between the same two adjacent fluid layers.

In symbols, shear stress $\tau \propto \frac{dv}{dy}$ (or) $\tau = -\eta \frac{dv}{dy}$

$$\gamma = \frac{F}{2l}$$

$$= \frac{0.04}{2 \times 0.16} = 0.13 \text{ N m}^{-1}$$

Example 3.12 Calculate the force required to pull a flat circular plate of radius 15 cm from the surface of a liquid which has a surface tension of 0.053 N m^{-1}.



Radius of circular plate $r = 15 \text{ cm} = 15 \times 10^{-2} \text{ m}$,

surface tension $\gamma = 0.053 \text{ N m}^{-1}$

circumference of the plate $l = 2\pi r$

$= 2 \times 3.142 \times 15 \times 10^{-2} = 94.26 \times 10^{-2} \text{ m}$

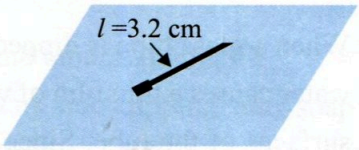
Since the circular plate has only one rim in contact with the liquid surface,

$$\gamma = \frac{F}{l}$$

$$0.053 = \frac{F}{94.26 \times 10^{-2}}$$

$$F = 0.0499 \text{ N}$$

Example 3.13 A needle has a length of 3.2 cm. When placed gently on the surface of the water in a glass, this needle will float if it is not too heavy. What is the weight of the heaviest needle? Assume that the surface tension of water is 0.073 N m^{-1}.



Surface tension of water $\gamma = 0.073 \text{ N m}^{-1}$, $l = 3.2 \text{ cm} = 3.2 \times 10^{-2} \text{ m}$,

Since the needle has two edges in contact with the liquid surface,

$$\gamma = \frac{F}{2l}$$

$$0.073 = \frac{F}{2 \times 3.2 \times 10^{-2}}$$

$$F = 4.67 \times 10^{-3} \text{ N}$$

This force is weight of the heaviest needle.

2. Capillary Action (Capillarity)

Capillary action is defined as the spontaneous flow of a liquid into a narrow tube or porous material. Capillary action is caused by the combination of cohesive forces of the liquid and the adhesive forces between the liquid and tube material. Cohesion and adhesion are two types of intermolecular forces. Cohesion refers to forces between molecules of same substance; whereas, adhesion refers to forces between molecules of different substances. These forces pull the liquid into the tube. In order for wicking to occur, a tube needs to be sufficiently small in diameter.

Capillarity (or) capillary action is also the rise or depression of a liquid in a narrow tube of small cross-sectional area, like the spaces between the fibres of a towel or the openings in a porous

Shear stress is the shearing force per unit area. That is $\tau = \frac{F}{A}$, where F is the shearing force acting between two adjacent layers of a fluid, η is the viscosity (coefficient of viscosity), A is the area of the fluid layer and $\frac{dv}{dy}$ is velocity gradient.

In Figure 3.8, $\Delta v = v_1 - v_2$, $\lim_{\Delta y \to 0} \frac{\Delta v}{\Delta y} = \frac{dv}{dy}$

Viscosity is defined as the ratio of shearing stress to the velocity gradient.

$$\eta = -\frac{F/A}{dv/dy} \tag{3.3}$$

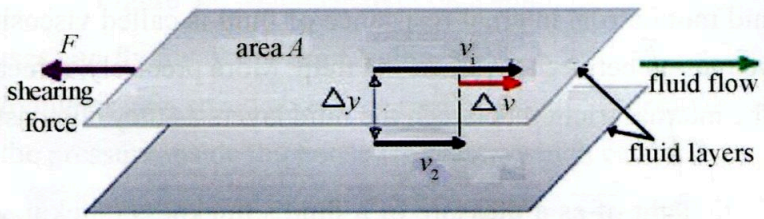


Figure 3.8 Force between two adjacent fluid layers

Determination of Kinematic Viscosity using Stokes' Law

Stokes' law, derived by the British scientist Sir George G. Stokes in 1851, expresses the drag force resisting the fall of small spherical objects through a fluid medium under the influence of gravity. It states that the drag force F acting upward in resistance to the fall of a spherical object in a viscous liquid is

$$F = 6\,\pi\,\eta\,r\,v \tag{3.4}$$

where η is the coefficient of viscosity of the liquid, r is the radius of the sphere, and v is the constant velocity of fall called the terminal velocity.

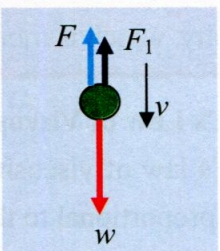
A simple way of measuring viscosity is that a small sphere ball is dropped through the fluid and time of fall and distance of fall are measured. From the distance and time of fall, the velocity can be determined.

Consider the fall of a sphere ball as shown in the diagram. The forces acting on the object are;

weight of the sphere ball $w = \frac{4}{3}\pi r^3 \rho_1 g$

upward thrust $F_1 = \frac{4}{3}\pi r^3 \rho_2 g$

drag force $F = 6\,\pi\,\eta\,r\,v$



where ρ_1 is the density of the sphere, ρ_2 is the density of the liquid, and g is the acceleration due to gravity.

At a constant velocity of fall (terminal velocity), the upward and downward forces are in balance.

Therefore, $$\frac{4}{3}\pi r^3 \rho_2 g + 6\pi\eta r v = \frac{4}{3}\pi r^3 \rho_1 g$$

$$6\pi\eta r v = \frac{4}{3}\pi r^3 (\rho_1 - \rho_2) g$$

$$\eta = \frac{2gr^2(\rho_1 - \rho_2)}{9v} \qquad (3.5)$$

Example 3.9 A metal plate of area 2.5×10^{-4} m^2 is placed on a 0.25×10^{-3} m thick layer of castor oil. If a force of 50 N is needed to move the plate with a velocity 3×10^{-2} m s^{-1}, calculate the coefficient of viscosity of castor oil.

$A = 2.5 \times 10^{-4}$ m^2 , $\Delta y = (0.25\times10^{-3} - 0) = 0.25\times10^{-3}$ m,

$F = -50$ N (the drag force is opposite to the direction of motion),

$\Delta v = (3 \times 10^{-2} - 0) = 3 \times 10^{-2}$ m s^{-1}

velocity gradient $$\frac{\Delta v}{\Delta y} = \frac{dv}{dy} = \frac{3\times10^{-2}}{0.25\times10^{-3}} = 120\ \text{m s}^{-1}$$

$$\eta = -\frac{F/A}{dv/dy}$$

$$= -\frac{(-50)}{2.5\times10^{-4}} \times \frac{1}{120} = 1.67 \times 10^3 \text{ Pa s}$$

Example 3.10 Determine the radius of a rain drop falling through air with terminal velocity 9 m s^{-1}. Viscosity of air is 1.8×10^{-5} Pa s, density of water is 1000 kg m^{-3}, density of air is 1.21 kg m^{-3}.

$v = 9$ m s^{-1}, $\eta = 1.8 \times 10^{-5}$ Pa s, density of water $\rho_1 = 1000$ kg m^{-3}, density of air $\rho_2 = 1.21$ kg m^{-3}

$$\eta = \frac{2gr^2(\rho_1 - \rho_2)}{9v}$$

$$r = \sqrt{\frac{9v\eta}{2g(\rho_1 - \rho_2)}}$$

$$= \sqrt{\frac{9\times9\times1.8\times10^{-5}}{2\times9.8\times(1000-1.21)}} = 2.7 \times 10^{-4} \text{ m}$$

Reviewed Exercise

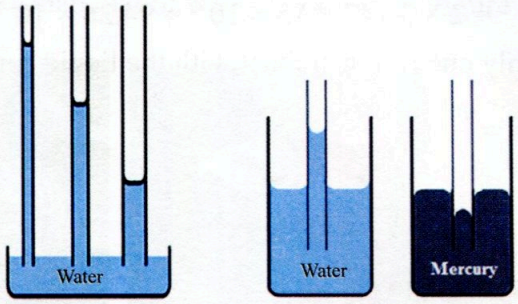
1. On which factors does the viscosity of a liquid depend?
2. How does temperature affect the viscosity of a fluid?

Key Words: intermolecular forces, intensive property, dynamic viscosity, kinematic viscosity

material. The rise of water underneath the ground is also the example of capillary action. Capillary action helps bringing water up into the roots.

When a glass tube with a small inside diameter is dipped into water, the water rises into the tube. For example, a glass tube with a bore of about 1 mm in diameter, the water will rise to a height of about 5 cm. The smaller the bore, the higher the water will rise as shown in Figure 3.10 (a).

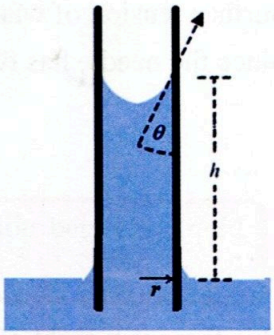
Liquids that rise in small bore tubes inserted into the liquid are said to wet the tube, whereas liquids that are depressed within thin tubes below the surface of the surrounding liquid do not wet the tube. Water is a liquid that wets glass capillary tubes; mercury is one that does not, as shown in Figure 3.10 (b). Capillarity does not occur for non-wetting liquids.



(a) (b)

Figure 3.10 Capillary action (a) rising water into the tube (b) comparing water and mercury

When a glass tube is dipped into water, the adhesion between glass and water causes a thin film of water to be drawn up over the inner and outer surfaces of the tube. Since the adhesive force is greater than cohesive force, water level rises until there is a sufficient mass of liquid for gravitational forces to overcome these intermolecular forces.



The actual amount of rise depends on the surface tension γ, the contact angle θ and the radius r of the tube. The height h to which the liquid can be lifted is given by,

$$h = \frac{2\gamma \cos\theta}{\rho g r} \qquad (3.7)$$

where ρ is density of liquid.

Example 3.14 When a capillary tube stands upright in a beaker of water, the water rises h in the tube. If the radius of the tube is 2 mm, contact angle is 30°, density of water is 1000 kg m^{-3} and surface tension is 0.072 N m^{-1} calculate h.

$\gamma = 0.072$ N m^{-1}, $\rho = 1000$ kg m^{-3}, $\theta = 30°$, $r = 2$ mm $= 2 \times 10^{-3}$ m

$$h = \frac{2\gamma\cos\theta}{\rho g r}$$

$$= \frac{2\times 0.072\times\cos 30^\circ}{1000\times 9.8\times 2\times 10^{-3}} = 0.0064\text{ m} = 0.64\text{ cm}$$

Example 3.15 At a certain temperature, water has a surface tension of 0.072 N m^{-1}. In a 3 mm diameter vertical tube if the liquid rises 6 mm above the liquid outside the tube, calculate the contact angle. Assume density of water ρ = 1000 kg m^{-3}.

$\gamma = 0.072$ N m^{-1}, $\rho = 1000$ kg m^{-3}, $h = 6\times 10^{-3}$ m,

$d = 3$ mm $= 3\times 10^{-3}$ m, $r = \frac{d}{2} = 1.5\times 10^{-3}$ m, $\theta = ?$,

$$h = \frac{2\gamma\cos\theta}{\rho g r}$$

$$\cos\theta = \frac{h\rho g r}{2\gamma}$$

$$= \frac{6\times 10^{-3}\times 1000\times 9.8\times 1.5\times 10^{-3}}{2\times 0.072} = 0.6125$$

$$\theta = 52.23^\circ$$

Reviewed Exercise

1. Water rises in a capillary tube, whereas mercury falls in the same tube. Explain.
2. A soap film is formed on a rectangular frame of length 7 cm side dipping into soap solution. The frame hangs from the arm of a balance. An extra weight of 0.4 g is to be placed in the opposite pan to balance the pull on the frame. Calculate the surface tension of soap solution.

Key Words: capillary tube, surface tension, viscosity, shearing force, terminal velocity

SUMMARY

If fluid particles are moving steady in smooth paths in layers, with each layer moving smoothly past the adjacent layers with no mixing; such a flow is called **laminar flow**.

If the flow or path of the fluid particles are irregular, their direction is always changing or whirling; this fluid movement is called **turbulent flow**.

Bernoulli's theorem states that the sum of pressure, kinetic energy per unit volume and potential energy per unit volume of an incompressible, non-viscous fluid in a streamlined laminar flow remains constant.

The resistance to fluid motion (or) internal resistance of fluid is called **viscosity**.

Surface tension is defined as the force per unit length that acts across any line on a surface tending to pull the surface apart.

EXERCISES

1. How are the flow rate and fluid velocity related?
2. A perfume bottle or atomizer sprays the fluid that is in the bottle. How does the fluid rise up in the vertical tube?
3. One hypodermic syringe contains medicine with density 1010 kg m^{-3}. The barrel of the syringe has a cross-sectional area 2.5×10^{-5} m^2, and the needle has a cross-sectional area 10^{-8} m^2 . The syringe is in horizontal position and injection is forced by 2 N acting on the plunger. Find the speed of the injection that enters into the patient. (The volume of syringe is 3 cc and injection time is 2 s)
4. In an adult, the radius of aorta is normally ~ 1.5 cm and blood moves through it at an average speed of 30 cm s^{-1}. If typical capillary has a radius of 5×10^{-6} m, and blood passes through them with a velocity of 0.1 cm s^{-1} approximately how many capillary are in the body?
5. An aeroplane's wings have a total surface area of 480 m^2. The pressure difference between the upper and lower surfaces of each wing is 6500 Pa. Calculate the lift created.
6. Derive expression for terminal velocity when a ball of radius r is dropped through a liquid of viscosity η and density ρ.
7. The shear stress at a point in a liquid is found to be 0.03 N m^{-2}. The velocity gradient at the point is 0.15 s^{-1}. What will be it's viscosity(in poise)? (1 Pa s = 10 poise)
8. A square plate 0.1 m side moves parallel to second plate with a velocity of 0.1 m s^{-1}, both plates being immersed in water. If the viscous force is 0.002 N and the coefficient of viscosity is 0.01 poise, what is the distance between the plates? (1 Pa s = 10 poise)
9. Assume that a spherical object is flowing through water. Viscosity of water is 0.001 Pa s, radius of spherical object is 2 mm and the velocity of the object at particular instant is 2 m s^{-1}. Find the drag force on the object due to the fluid. Assume that Stokes' law is valid.
10. Which force causes capillary action rise?
11. How does the temperature affect capillary action?
12. Identify which property of water allows water to move against gravity.
    A. Capillary action allows water to move against gravity.
    B. Surface tension allows water to move against gravity.
    C. Cohesion allows water to move against gravity.
13. Due to capillary action, a liquid will rise in a tube, if the angle of contact is
    A. acute B. obtuse C. 90°
14. What will happen to the rise of liquid in a capillary tube if the radius of the tube is (i) increased, (ii) decreased?
15. Calculate the height to which water will rise in a capillary tube of diameter 1 mm. The surface tension of water is 0.072 N m^{-1}, density of water is 1000 kg m^{-3}. (Assume angle of contact is 0°.)

CONCEPT MAP

FLUID DYNAMICS

Equation of Continuity

$A_1 v_1 = A_2 v_2$

Bernoulli's Equation

$p + \frac{1}{2}\rho v^2 + \rho g h = \text{constant}$

Viscosity

- Newton's law of viscosity

$\eta = -\frac{F/A}{dv/dy}$

- kinematic viscosity using Stokes' law

$F = 6\pi\eta r v$

Manifestations of Bernoulli's theorem

1. lift on the wings of an aeroplane
2. insect sprayer
3. Magnus effect on a spinning ball
4. roof of houses can fly away during cyclones

SURFACE TENSION AND CAPILLARITY

Surface Tension

$\gamma = \frac{F}{l}$

Capillary Action

$h = \frac{2\gamma\cos\theta}{\rho g r}$

CHAPTER 4

SOURCES OF ENERGY AND ENVIRONMENTAL IMPACTS

Energy resources are all forms of fuel used in the modern world, either for heating, generation of electrical energy, or for other forms of energy conversion processes. The different types of energy sources cause more or less impacts on the environment. Renewable energy sources (non-conventional energy sources) do not pollute the environment and natural surrounding and require less expenditure. In this chapter, we will learn about different sources of energy and their environmental impacts.

Learning Outcomes

It is expected that students will

- deduce the knowledge of energy demand and crisis.
- recognize the different types of energy sources, their fundamental characteristics and usefulness.
- identify the challenges and problems associated with the use of various energy sources.
- understand the extent of environmental impact of the non-renewable and renewable sources of energy.

4.1 ENERGY DEMAND AND CRISIS

Energy is the basic necessity for life. Energy demand is the term used to describe the consumption of energy by human activities. It drives the whole energy system, influencing the total amount of energy used; the locations and types of fuel used in the energy supply system and the characteristics of the technologies that consume energy. Energy demand depends on different socioeconomic factors such as population, urbanization, industrialization, net capital income and development of technologies, etc.

Energy demand forecasting is crucial for ensuring future economic growth and environmental security. There has been an enormous increase in the global demand for energy in recent years as a result of industrial development and population growth. Supply of energy is, therefore, far less than the actual demand. Because the population and new technology are always expanding demand for energy is expected to increase year by year.

The energy crisis is the concern that the world's demands on the limited natural resources that are used to power industrial society are diminishing as the demand rises. It is a situation in which a nation suffers from a disruption of energy supplies connected by increasing energy prices that threaten economic and national security.

With ninety percent of the world's oil reserves already discovered; people need to find new ways

to make energy. Since the non-renewable resources of the energy will not last long and cannot be regained, the world is going to face the problem of energy crisis in near future. Renewable energy has huge potential to provide solution to increase energy crisis and it is the key factor to the future of energy, food and economic security.

Example 4.1 At present day, the rate of formation of fossil fuels (coal, petroleum, natural gas) is estimated to be about 4 billion kW (4×10^9 kW) and the world's consumption of energy is of the order of 10^{17} Btu/year. Show that the rate of consumption of fossil fuel is of the same order of magnitude as the rate of their formation and these energy sources are therefore non-renewable.

Today's world consumption of energy is of the order of 10^{17} Btu/year. This number need to be converted to kW.

Using 1 Btu = 1055 J and 1 year = 3.15×10^7 s

$$10^{17} \text{ Btu/year} = \frac{10^{17} \times 1055}{1 \times 3.15 \times 10^7} \text{ J s}^{-1} = 3.35\times10^{12} \text{ W} = 3.35 \times 10^9 \text{ kW}$$

Hence, the rate of consumption of fossil fuel is of the same order of magnitude as the rate of their formation and these energy sources are therefore non-renewable.

Reviewed Exercise

- Why does the energy demand increase year by year?

Key Words: energy demand, crisis, consume energy

4.2 SOURCES OF ENERGY

In our daily life we use different forms of energy such as heat energy, light energy, mechanical energy, electrical energy, chemical energy and sound energy. The most common forms of energy are heat, light and electricity. All the sources of energy can be divided into two categories: renewable sources and non-renewable sources of energy.

A non-renewable resource is a natural substance that is not replenished with the speed at which it is consumed. It is a finite resource. Non-renewable energy resources include coal, natural gas, crude oil, and nuclear energy. They occur in limited and exhaustible quantities and cannot be regenerated in a short period of time or used again and again.

Renewable energy sources (also called non-conventional energy sources) are continuously replenished by natural processes. It is the energy generated by using wind, solar, hydro, tides, etc. They cannot be exhausted easily and can be generated at a constant rate for their use over and over again. They can be produced or generated through natural processes, at a rate greater than or equal to the rate of their consumption.

Non-renewable energy sources are available only in limited supply. Examples include fossil fuels (like coal, oil, natural gases, etc.) and nuclear energy. It takes a long time (often millions of years) for these resources to be replenished.

Renewable resources are replenished naturally and over short periods of time, so are in a virtually unlimited supply. Examples include the solar energy, hydro energy, wind energy and geothermal energy.

Figure 4.1 Energy resources

Reviewed Exercise

- What is the main source of energy?

Key Words: renewable sources of energy, non-renewable sources of energy

4.3 NON-RENEWABLE SOURCES OF ENERGY

1. Fossil Fuels

Fossil fuels, such as coal, oil and natural gas, are important non-renewable sources of energy. Carbon is the main constituent of these fossil fuels. A large amount of chemical energy is stored in the fossil fuels. This stored chemical energy is converted into various other forms of energy such as heat, light and mechanical energy. Fossil fuels are excellent sources of energy for our transportation needs. For example, the fossil fuel industry extracts crude oil from the ground and converts it to gasoline. Fossil fuel liquids also are refined into petrochemical products that are used as ingredients in the manufacture of literally hundreds of products from plastics and polyurethane to solvents.

Coal is formed in a way similar to the other fossil fuels, though it goes through a different process called coalification. It is made of decomposed plant matter in conditions of high temperature and pressure, and it takes a relatively shorter amount of time to form. Coal is not a uniform substance; its composition varies from deposit to deposit.

Natural gas is another major source of the energy in our country. Oil and gas fields have been found some places on the earth. These fields always contain some gases, but this natural gas (methane) does not take nearly as long to form. Natural gas is also found in independent deposits within the ground as well as from other sources too. Methane is a common gas found in swamps and is also the by-product of animals' digestive system.

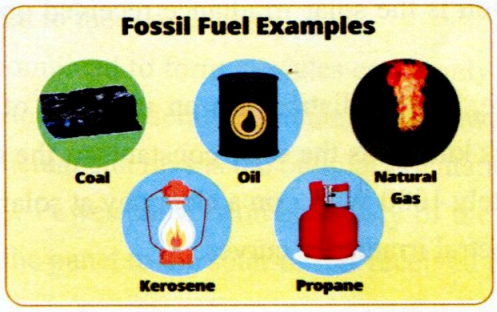


Figure 4.2 Examples for fossil fuel

2. Nuclear Energy

Nuclear energy is non-renewable as the uranium fuels used are consumed in the nuclear fission reaction and hence are non-replenishable. Nevertheless, nuclear energy has many uses:

- energy produced in a nuclear reactor can be harnessed to produce electricity,
- nuclear energy is also being used to power submarines and ship. Vessels driven by nuclear energy can sail for long periods without having to refuel,
- radioisotopes obtained as by-products in nuclear reactions are used in medicine, agriculture and research.

Reviewed Exercise

- What are the main characteristics of fossil fuels?

Key Words: fossil fuels, petrochemical products, nuclear energy

4.4 RENEWABLE SOURCES OF ENERGY

There are several alternative and renewable sources of energy which are not only environment friendly but can also be available in abundance.

The major types of renewable energy sources are: (i) solar, (ii) wind, (iii) hydropower, (iv) ocean, (v) geothermal and (vi) biomass.

1. Solar Energy

Sun is one of the most powerful renewable sources of energy. Solar energy is the radiant light and heat from the sun that has been harnessed by humans since ancient times using a range of

ever-evolving technologies. Earth's atmosphere and clouds absorb or scatter as much as 54 % of the incoming sunlight. The sunlight that reaches the ground consists of visible light, infrared, ultraviolet and other forms of electromagnetic radiation.

The total amount of radiation energy passing through per unit area and per unit time is called irradiance and is measured in W m^{-2} or kW m^{-2}. Another commonly used term for irradiance of sunlight is insolation. Insolation is the solar irradiance received on a horizontal surface such as ground.

Above the earth's atmosphere, solar radiation has an intensity of approximately 1350 W m^{-2} (429 Btu h^{-1} ft^{-2}). This value is known as the solar constant. At the earth's surface solar radiation has an intensity of approximately 1000 W m^{-2} on a clear day at solar noon in the summer months. Figure 4.3 shows the solar spectral irradiance curves.

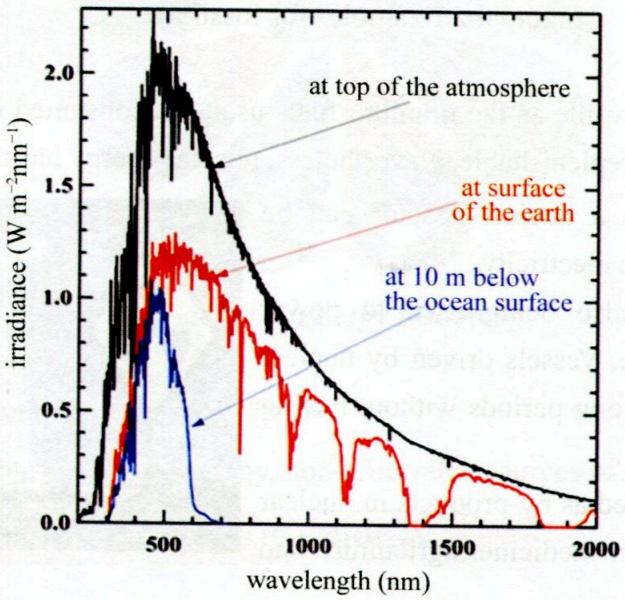


Figure 4.3 The solar irradiance curves at the top of the atmosphere and at surface of the earth

Solar Technologies

There are three main ways to harness solar energy: photovoltaic solar cell, solar heating, and concentrated solar power (CSP).

Photovoltaic Solar Cell

Conversion of light energy into electrical energy is based on a phenomenon called photovoltaic effect. When semiconductor materials are exposed to light, some of the photons of light are absorbed by the semiconductor crystal which causes a significant number of free electrons in the crystal. This is the basic reason for producing electricity due to photovoltaic effect.

A solar cell (also known as a photovoltaic cell) is defined as an electrical device that converts light energy into electrical energy through the photovoltaic effect. A solar cell is basically a p-n junction diode. When light is incident on the photovoltaic cell, the light photons can easily reach the p-n junction. The light energy, in the form of photons, supplies sufficient energy to the junction to create a number of electron-hole pairs as shown in Figure 4.4. As a result, a current flow in the external circuit connected to the p-n junction.

Individual solar cells can be combined to form modules commonly known as solar panels. When combined into a large solar panel, considerable amounts of renewable energy can be generated. Achieving the perfect perpendicular sun rays is the optimal orientation so that solar panels receive direct light throughout the day. Power conversion efficiency of the solar panel is the ratio of the electrical power delivered by the panel to the solar power received by it.

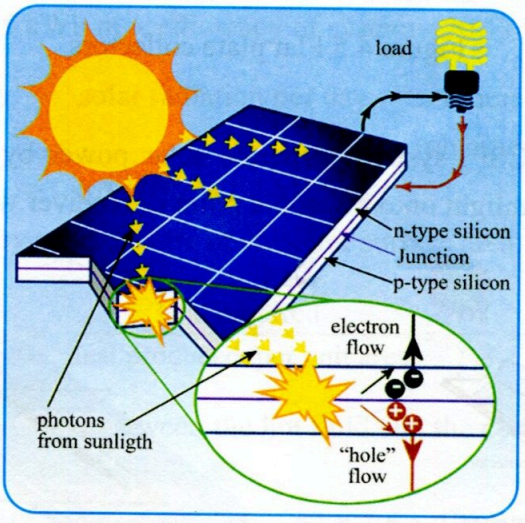


Figure 4.4 Photovoltaic solar cell

Solar Heating

Solar heating can also play an important role in reducing the impact of global warming. Solar water heating systems installed in homes are comprised of three main elements: the solar collector, insulated piping, and a hot water storage tank. The solar collector gathers the heat from solar radiation and transfers the heat to water. This heated water flows out of the collector to a hot water tank, and is used as necessary.

Flat plate collectors are the most common type of collector. Flat plate collectors have copper pipes attached to an absorber plate contained in an insulated box that is covered with a tempered glass or polymer cover plate.

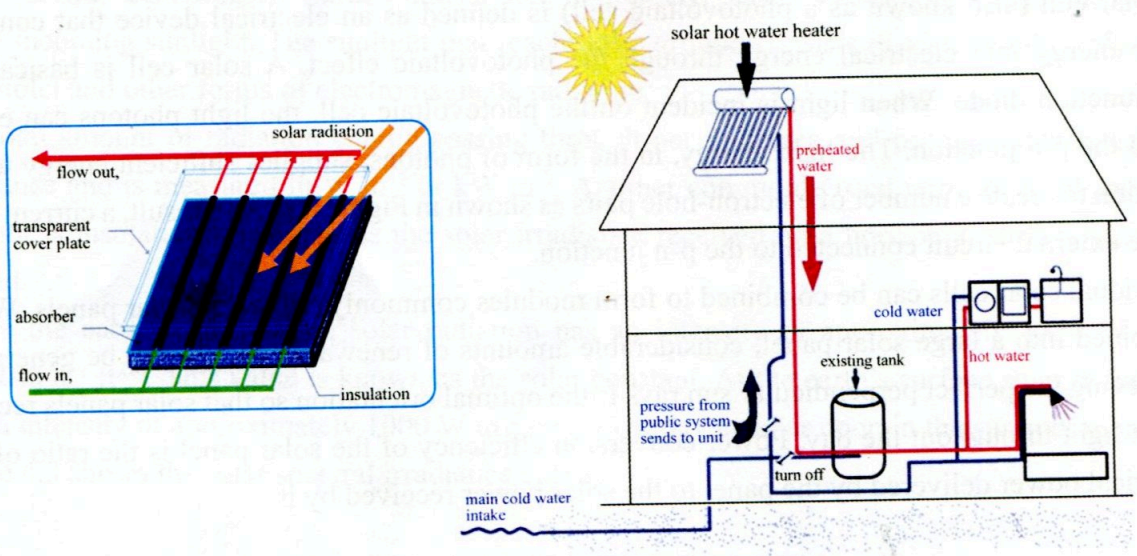


Figure 4.5 Flat plate collectors

Concentrated Solar Power

Concentrated solar power (CSP) systems generate solar power by using mirrors or lenses to concentrate a large area of sunlight onto a specific point of receiver which converts into heat.

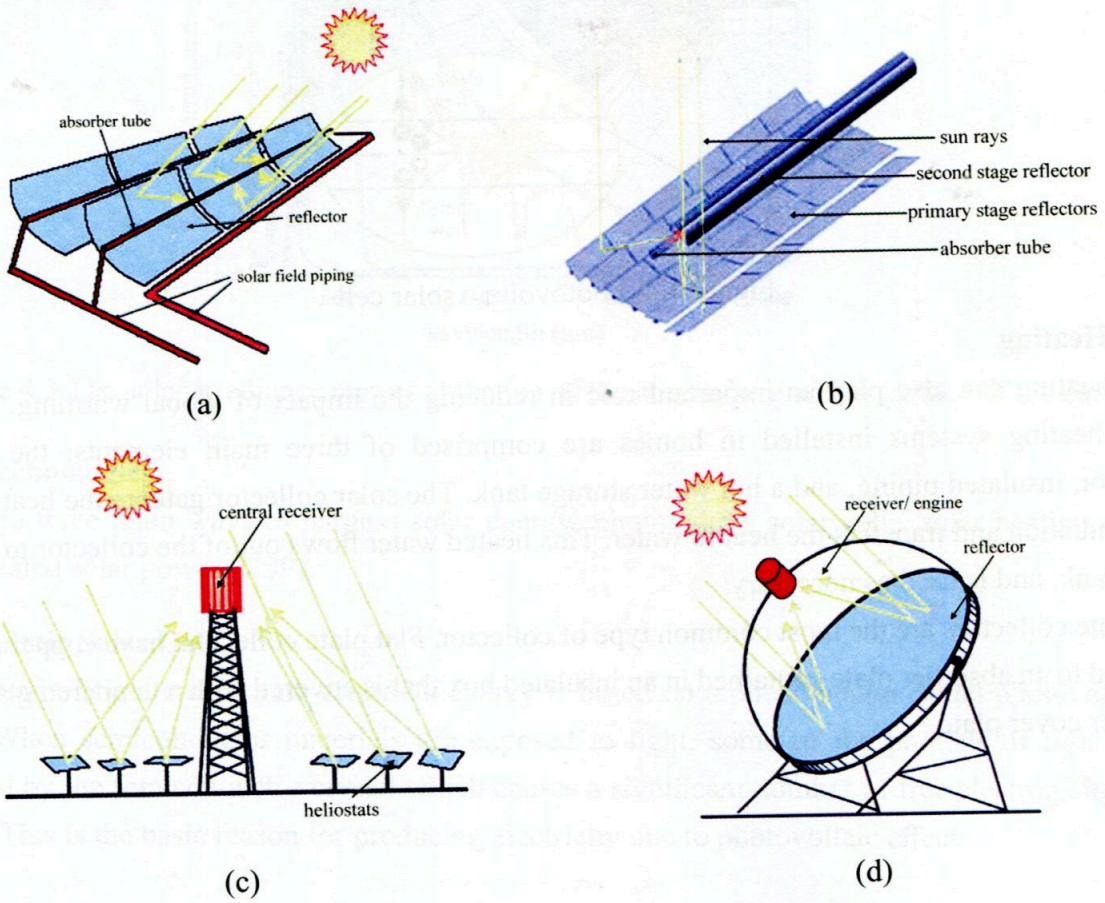


Figure 4.6 Four types of concentrated solar power (CSP) systems

The heat is then used to create steam, which drives a turbine to generate electrical power. Therefore CSP plants consist of two parts: one that collects solar energy and converts it to heat, and another that converts the heat energy to electricity.

There are four types of CSP technologies: (a) Parabolic trough systems, (b) Linear fresnel systems, (c) Power tower systems and (d) Parabolic dish systems. Figure 4.6 (a), (b), (c) and (d) show the four types of CSP technologies.

Example 4.2 A home requires 62 kWh of heat on a winter day to maintain a constant indoor temperature of 20 °C. Assume that the temperature of hot water outlet is 60 °C. (i) How much collector surface area does it need for a solar heating system that has a 20 % efficiency? (ii) How large does the storage tank have to be to provide this much energy? (The average solar radiation per day in winter is about 6.5 kWh m^{-2})

Solar radiation per day = 6.5 kWh m^{-2}, efficiency of system= 20 % = 0.2

(i) Thermal energy per day = solar radiation per day × efficiency of system (obtained using collector)

$$= 6.5 \times 0.2 = 1.3 \text{ kWh m}^{-2}$$

This means that for every square metre of collector surface area, 1.3 kWh of heat are produced every day.

$$\text{Collector surface area} = \frac{\text{heat required}}{\text{heat produced per unit area}} = \frac{62}{1.3} = 48 \text{ m}^2$$

(ii) Temperature difference is that between the hot fluid and the cold water going into the storage tank = 60 – 20 = 40 °C = 40 K

Since 1 kWh = 3.6×10^6 J, $\Delta Q = 62$ kWh = $62 \times 3.6 \times 10^6$ J

Specific heat capacity of water $c = 4184$ J kg^{-1} K^{-1}

$$\Delta Q = m c \Delta T$$

$$m = \frac{\Delta Q}{c \Delta T}$$

$$= \frac{62 \times 3.6 \times 10^6}{4184 \times 40} = 1333.6 \text{ kg}$$

This is equivalent to a volume of 1334 liters (or about 352.4 gallons).

Reviewed Exercise

1. The band gap of GaAs is 1.4 eV. Calculate the minimum wavelength of light for photovoltaic generation in a GaAs solar cell.
2. A 100 ft × 50 ft building has a flat roof. What is the average solar energy received by the roof in a month? (The solar radiation per day in winter is about 1600 Btu ft^{-2})

3. Calculate the solar energy received by a standard hot water collector of dimensions 1 m × 2 m, over one hour at around noon, if the irradiance stays fairly constant at about 800 W m^{-2}.

Key Words: irradiance, insolation, photovoltaic cell, concentrated solar power

2. Wind Energy

Wind energy is the energy generated or produced by harnessing the power of the wind. Like solar power, harnessing the wind is highly dependent on weather and location. However, it is one of the oldest and cleanest forms of energy and the most developed of the renewable energy sources. Both windmills and wind turbines harness wind energy and put it to practical use. Windmills convert wind energy directly into mechanical energy for such tasks as milling grain and pumping water. A wind turbine converts wind energy into electricity, which can then be used to power electrical equipment, stored in batteries or transmitted over power lines.

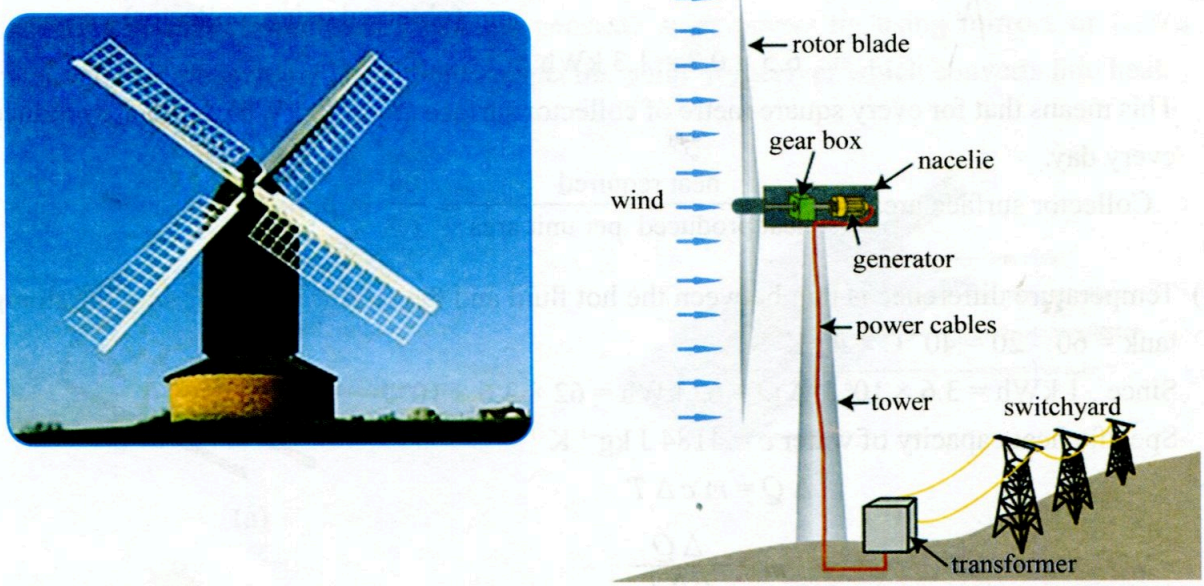


Figure 4.7 Windmill and wind turbines

In a wind turbine, wind first hits the blades of the turbine, causing them to rotate and turn the turbine connected to them. That changes the kinetic energy to rotational energy, by rotating a shaft which is connected to a generator, and thereby producing electrical energy through electromagnetism. Wind farms consist of many individual wind turbines, which are connected to the electric power transmission network.

The amount of power that can be harvested from wind depends on the size of the turbine and the length of its blades. Wind energy is the kinetic energy of air in motion. Total wind energy flowing through an imaginary surface with area A during the time t is calculated as follows,

$$E = \frac{1}{2} m v^2 = \frac{1}{2} \rho V v^2 = \frac{1}{2} \rho A v t v^2 = \frac{1}{2} \rho A t v^3$$

where ρ is the density of air, v is the wind speed, $A\ v\ t$ is the volume V of air passing through area A (which is considered perpendicular to the direction of the wind) in time t.

Power is energy per unit time, so the wind power incident on A is,

$$P = \frac{1}{2} \rho A v^3 \qquad (4.1)$$

where A = the rotor area of a wind turbine.

The output is proportional to the dimensions of the rotor and to the cube of the wind speed. Theoretically, when wind speed doubles, wind power potential increases by a factor of eight.

Wind turbines, like windmills, are mounted on a tower to capture the most energy. At 100 feet (30 meters) or more above ground, they can take advantage of the faster and less turbulent wind. Turbines catch the wind energy with their propeller-like blades. Usually, two or three blades are mounted on a shaft to form a rotor.

Example 4.3 Determine the input power of the wind if the wind speed is 20 m s^{-1} and blade length is 20 m. Density of air is 1.3 kg m^{-3}.

Radius of blade of turbine r = 20 m, wind speed = 20 m s^{-1}, density of air = 1.3 kg m^{-3}

Area of a wind turbine $A = \pi r^2 = 3.142 \times (20)^2 = 1256.8\ m^2$

Wind power $P = \frac{1}{2} \rho A v^3$

$= \frac{1}{2} \times 1.3 \times 1256.8 \times (20)^3$

$= 6.535 \times 10^6\ W = 6.535\ MW$

Reviewed Exercise

- How does wind power generate electricity?

Key Words: windmill, wind turbine

3. Hydroelectric Energy

The flowing water and water stored in huge dams is also a very important source of energy which is known as hydroelectric energy. Hydropower, or hydroelectric power, is one of the oldest and largest sources of renewable energy. The source of power in a hydroelectric plant is the motion of water. It is an affordable source of electricity that costs less than most of other sources of energy. However, overdevelopment and unrestricted harnessing of water power can have a devastating effect on the local environment and habitation areas. There are three types of hydropower facilities: impoundment, diversion, and pumped storage.

Hydroelectric energy, also called hydroelectric power or hydroelectricity, is a form of energy that harnesses the power of water in motion such as water flowing over a waterfall to generate electricity.

Impoundment Facility

The most common type of hydroelectric power plant is an impoundment facility. It is typically a large hydropower system, uses a dam to store river water in a reservoir. Water is collected or stored at a higher elevation and gravity pulls the water downward through large pipes or tunnels (penstocks) to a lower elevation; the difference in these two elevations is known as the head. At the end of its passage down the pipes, the falling water causes turbines to rotate. The turbines in turn drive generators, which convert mechanical energy into electricity. Transformers are then used to convert the alternating voltage of the generators to a higher voltage suitable for long-distance transmission.

A penstock is a closed conduit that channels the flow of water to turbines with water flow regulated by gates and valves.

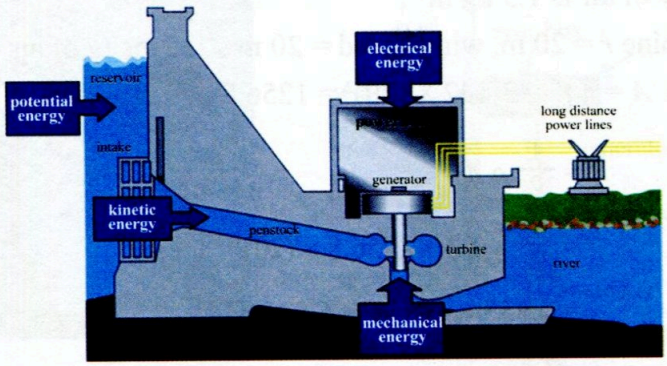
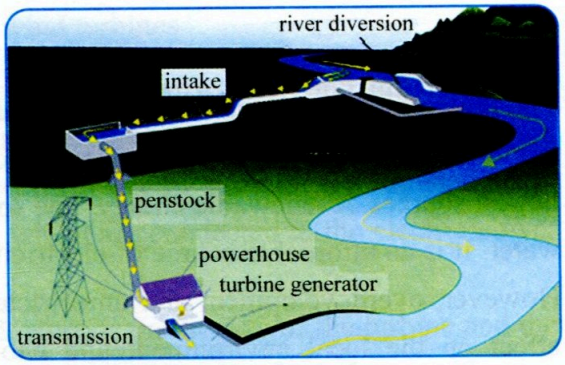


Figure 4.8 Hydroelectric power plant

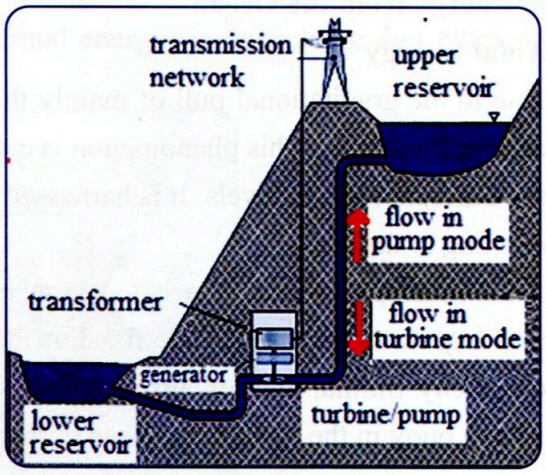
Diversion Facility (or) Run-of-river Facility

A diversion, sometimes called a run-of-river facility, channels a portion of a river through a canal or a penstock to utilize the natural decline of the river bed elevation to produce energy. A diversion may not require the use of a dam.



Pumped Storage Facility

Another type of hydropower is pumped storage hydropower (PSH). A PSH facility is able to store the electricity generated by other power sources. When the demand for electricity is low, a PSH facility stores energy by pumping water from the lower reservoir to an upper reservoir at a higher elevation. During periods of high electrical demand, the water is released back to the lower reservoir and turns a turbine, generating electricity.



Available Hydroelectric Power

A hydropower resource can be evaluated by its available power. The power available from falling water can be calculated from the flow rate and density of water, the height of fall, and the local acceleration due to gravity.

$$P = \eta \rho \dot{V} g h$$

where P is useful power output (W), η is efficiency of the turbine (dimensionless), ρ is density of water, h is difference in height between the outlet and inlet and $\dot{V} = \frac{V}{t}$ = volumetric flow rate. It is measured in $m^3\ s^{-1}$.

Most conventional hydroelectric plants include most common parts, (i) dam (raises the water level of the river and controls flow of water) (ii) turbine (a water turbine converts the kinetic energy of falling water into mechanical energy) (iii) generator (connected to the turbine by shafts and converts the mechanical energy from the turbine to electrical energy) (iv) transformer (v) transmission system (transport electricity from the hydropower station to consumers).

Example 4.4 At a hydroelectric power plant, the water pressure head is at a height of 300 m and the water flow available is 100 $m^3\ s^{-1}$. If the turbine generator efficiency is 60 %, estimate the electric power available from the plant. ($g = 9.8\ m\ s^{-2}$, $\rho = 1000\ kg\ m^{-3}$)

$h = 300$ m, Volume of water flowing per second $\dot{V} = 100\ m^3\ s^{-1}$, $\eta = 60\ \% = 0.6$

Available electric power $P = \eta \rho \dot{V} g h$

$= 0.6 \times 1000 \times 100 \times 9.8 \times 300$

$= 17.64 \times 10^7\ W = 176.4\ MW$

Reviewed Exercise

- The hydroturbine receives water from a reservoir at an elevation of 100 m above it. What is the minimum water flow in kg s^{-1} to produce a steady turbine output of 50 MW if the turbine generator efficiency is 75 %?

Key Words: dams, hydroelectric, hydropower, reservoir, pumped storage, efficiency

4. Energy from the Ocean

Tidal Energy

Due to the gravitational pull of mainly the moon on the spinning earth, the level of water in the sea rises and falls. This phenomenon is called high and low tides. The tidal energy is produced by the difference in sea levels. It is harnessed by constructing a dam (a tidal barrage) across a narrow opening to the sea.

A tidal barrage works similarly to a large hydropower reservoir dam but it is placed at the entrance to a bay or estuary. A turbine fixed at the opening of the tidal barrage converts tidal energy to electricity. Similar to river current technologies, turbines anchored to the ocean floor or suspended from a buoy in the path of an ocean current could also be used to generate power.

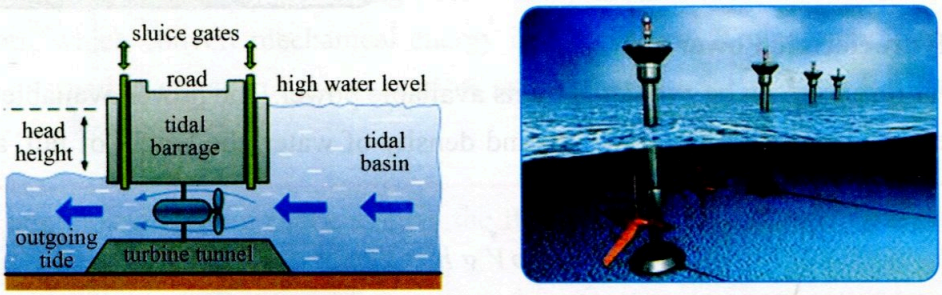


Figure 4.9 Ocean tidal power plant

Wave Energy

The kinetic energy of huge waves near the seashore can be trapped to generate electricity. The waves are generated by strong winds blowing across the sea. A variety of technologies are being tested to convert wave energy into electricity. Most systems capture energy on the surface of waves or use pressure differences just below the surface.

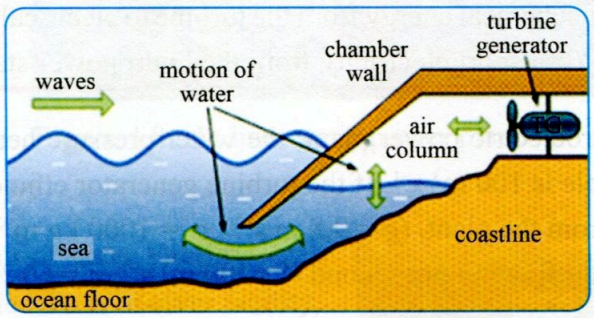


Figure 4.10 Conversion of wave energy into electricity

Ocean Thermal Energy

The water at the surface of the sea or ocean is heated by the sun while the water in deeper sections is relatively cold. This difference in temperature is exploited to obtain energy in ocean-thermal-energy conversion plants. These plants can operate if the temperature difference between the water at the surface and water at depths up to 2 km is 20 K (20 °C) or more. The warm surface-water is used to boil a volatile liquid (e.g. ammonia, boiling point ~ −33.34 °C). The vapours of the liquid (ammonia) are then used to run the turbine of generator. The cold water from

the depth of the ocean is pumped up and condense vapour again to liquid. The energy potential from the sea (tidal energy, wave energy and ocean thermal energy) is quite large, but efficient commercial exploitation is difficult.

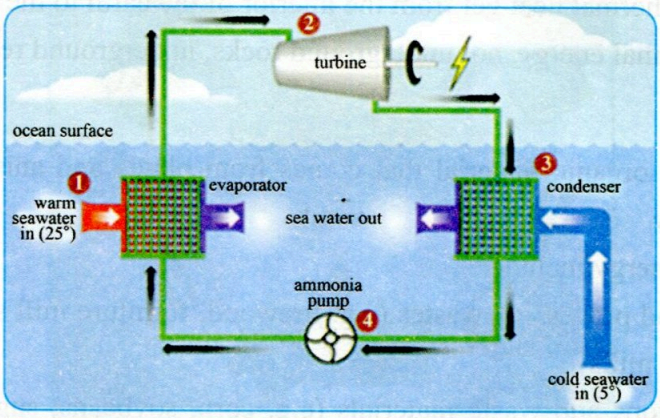


Figure 4.11 Conversion of ocean thermal energy into electricity

Reviewed Exercise

1. Why are wind and solar energies more popular than ocean energy?
2. What type of working fluids do closed-cycle ocean thermal energy conversion systems use?

Key Words: ocean thermal energy, tidal energy, wave energy

5. Geothermal Energy

Geothermal energy is a type of renewable energy taken from the Earth's core (the internal heat of the earth) and it is one of the oldest types of natural sources of heat. However, it is limited to regions of the world where there are hot underground rocks. Thermal energy from magma can be used to generate electricity. Magma is extremely hot liquid and semi-liquid rock located under Earth's surface. Wells of up to a mile deep or more are drilled into underground reservoirs to tap into the geothermal resources. These resources can be exploited from naturally occurring heat, rock and water permeability. Cooled water is pumped into heated rocks where it is turned into steam, which then rises up to drive the turbines and generate electricity. Condensed water is cooled and the process is repeated.

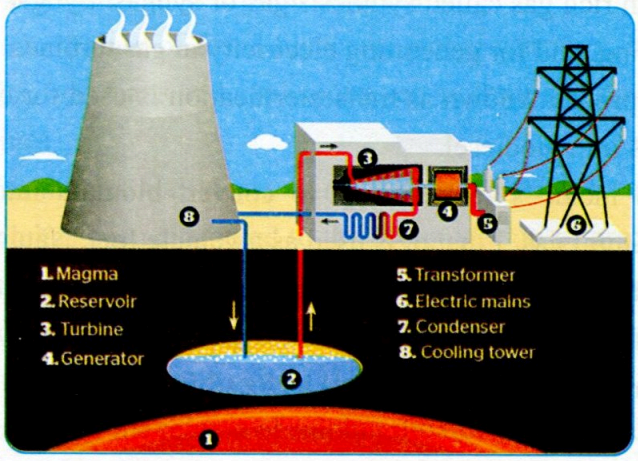


Figure 4.12 Electricity generated from geothermal energy

Reviewed Exercise

1. What is geothermal energy?
2. How does geothermal heat get from the interior of the earth to the surface?

Key Words: geothermal energy, hot underground rocks, underground reservoirs, magma

6. Biomass

Biomass is renewable organic material that comes from plants and animals. Biomass contains stored chemical energy.

Biomass sources for energy include:

- Wood and wood processing wastes (e.g. firewood, furniture mill sawdust and waste from pulp and paper mills).
- Agricultural crops and waste materials (e.g. corn, soybeans, sugar cane, woody plants, algae, rice husk and food processing residues).
- Biogenic materials in municipal solid waste (e.g. paper, cotton, wool products, food and wood wastes).
- Animal manure and human sewage.

Biomass is converted to energy through various processes, including: direct combustion (burning) to produce heat, thermochemical conversion to produce solid, gaseous, and liquid fuels, chemical conversion to produce liquid fuels, biological conversion to produce liquid and gaseous fuels.

Direct combustion is the most common method for converting biomass to useful energy. All biomass can be burned directly for heating buildings and water, for industrial heat process, and for generating electricity in steam turbines.

Thermochemical conversion of biomass includes pyrolysis and gasification. Both are thermal decomposition processes in which biomass feedstock materials are heated in closed, pressurized vessels called gassifiers at high temperatures. Biomass pyrolysis produces fuels such as charcoal, bio-oil, renewable diesel, methane, and hydrogen. Biomass gasification produces a carbon monoxide and hydrogen rich gas called synthesis gas or syngas. Syngas can be used as a fuel for diesel engines, for heating, and for generating electricity in gas turbines.

Chemical conversion process known as transesterification is used for converting vegetable oils, animal fats, and greases into biodiesel.

Biological conversion includes fermentation to convert biomass into ethanol and anaerobic digestion to produce natural gas. Ethanol is used as a vehicle fuel. Natural gas, also called biogas or biomethane, is produced in anaerobic digesters at sewage treatment plants and at dairy and livestock operations. It also forms in and may be captured from solid waste landfills. Properly treated renewable natural gas has the same uses as fossil fuel natural gas.

Figure 4.13 Biomass sources for energy

Reviewed Exercise

1. What is biomass?
2. Give the types of biomass energy sources.

Key Words: Biomass, direct combustion, biological conversion, chemical conversion

4.5 ENVIRONMENTAL IMPACTS AND CONSEQUENCES

Apart from the problem of energy resources getting depleted, energy harnessing and utilization cause an immense amount of environmental damage. There are environmental problems associated with the use of almost all forms of energy. The environmental problems directly related to energy production and consumption include air pollution, climate change, water pollution, thermal pollution, and solid waste disposal. It is important to understand the environmental impacts associated with producing electrical power from both non-renewable and renewable sources.

1. Environmental Impacts of Conventional Sources

The problems that relate to the use of fossil fuels are: global warming, acid rains, dangers posed by leaded fuels, oil spills, gas leaks and explosions, air pollution and water pollution caused by poorly managed coal mines.

Global warming refers to the gradual increase in the average temperature of the earth's surface and its atmosphere which has been attributed to the accumulation of greenhouse gases. The main greenhouse gases are carbon dioxide (CO_2), methane (CH_4), water vapour, oxides of nitrogen (NOx) and chlorofluorocarbons (CFCs). CFCs also destroy the earth's protective ozone layer which shields the earth from harmful ultraviolet (UV) rays from the sun.

Acid rains are caused mainly by the release of sulphur dioxide (SO_2) and oxides of nitrogen (NOx) when fossil fuels burn. The oxides combine with water vapour in the air to form acids, which

return to the ground as acid rain. Because carbonic acid is a relatively weak acid, the ability of carbon dioxide alone to generate true acid rain is very limited. It is important to note that acidified clouds could travel great distances before releasing the acid rain. The problems posed by acid rains include corrosion of the built environment, soil degradation, water pollution and depletion of forests.

The oil industry adds lead (Pb) to petrol (gasoline) to help engines run more smoothly. Vehicles that burn leaded gasoline pour out leaded fumes that contaminate the air. The World Health Organisation (WHO) has established that smoke from the combustion of leaded fuels in vehicles causes cancer and high blood pressure in adults; and in children, it impairs mental development, reduces intelligence thus hindering learning ability and causes behavioural disorders.

The leakage of fuel oil from storage vessels, oil tankers, pipelines and tanker trucks is unavoidable. Oil spills seriously damage the land, vegetation, and water bodies, including the oceans. In addition, spilled oil can damage the ecosystem and its habitats.

Gas leaks and explosions sometimes accompany the harnessing and utilization of fossil fuels especially in the coal mines and storage plants. The explosions are sometimes accompanied by fire outbreaks.

Excavated areas that have been strip mined for coal but are not filled and revegetated cause water pollution as surface water runoff from the mined area can flush sediments and sulphur-bearing compounds into nearby streams and rivers. This could endanger human life, plant and wildlife communities.

Emissions from vehicles, thermal power plants and factories contain unburned hydrocarbons, particulates, carbon dioxide, carbon monoxide and oxides of nitrogen and sulphur that contribute to the lowering of the quality of air. These substances in the air could irritate the eyes, throat and the lungs.

Figure 4.14 Environmental impacts of conventional sources

2. Environmental Impacts of Non-Conventional Sources

Environmental Impacts of Solar Power

Photovoltaics is now a proven technology which is inherently safe, as opposed to some dangerous

electricity generating technologies. Over its estimated life a photovoltaic module will produce much more electricity than was used in its production. A solar module will prevent the emission of CO_2 produced by other conventional energy sources. Photovoltaic systems make no noise and cause no pollution while in operation. PV cell technologies have relatively lower environmental risks compared to other types of electric sources. However, chemicals used in PV cells in the manufacturing facility, the installation site, and the disposal or recycling facility could be released to air, surface water, and groundwater.

Environmental Impacts of Wind Power

A wind farm, when installed on agricultural land, has one of the lowest environmental impacts of all energy sources. It occupies less land area per kilowatt hour (kWh) of electricity generated than any other energy conversion system. Greenhouse gas emissions and air pollution produced by its construction are very tiny and declining.

Modern wind turbine designs have significantly reduced the noise from turbines. Low frequency sound and infrasound (i.e. usually beneath the threshold of human hearing) are everywhere in the environment. Modern turbine designs which locate the blades upwind instead of downwind have significantly reduced the level of infrasound. Wind turbines may create shadow flicker on nearby residences when the sun passes behind the turbine. However, this can easily be avoided by locating the wind farm to avoid unacceptable shadow flicker, or turning the turbine off for a few minutes of the day when the sun is at the angle that causes flicker.

Environmental Impacts of Hydroelectric Power

Although hydropower has no air quality impacts, construction and operation of hydropower dams can significantly affect natural river systems as well as fish and wildlife populations. It has major effect on river transportation, agriculture, fertile lands, etc. Assessment of the environmental impacts of a specific hydropower facility requires case-by-case review.

Negative impacts of dams are as follows:

- in flat basins large dams cause flooding of large tracts of land, destroying local animals and habitats,
- villages and people have to be displaced causing change in life style,
- large amounts of plant life are submerged and decay anaerobically,
- the migratory pattern of river animals are affected,
- dams restrict sediments that are responsible for the fertile lands downstream,
- large dams are breeding grounds for mosquitoes and cause the spread of disease,
- dams serve as a heat sink, water in dans is hotter than the normal river water and when this warm water released into the river downstream can affect animal life.

Environmental Impacts of Geothermal Energy

Geothermal power is a relatively benign (less harmful) source of energy. For the most part, the impacts of development are positive. Worldwide geothermal energy utilization increases yearly because it is an attractive alternative to burning fossil fuels. Electricity generation from geothermal resources involves much lower greenhouse gas emission rates than that from fossil fuels.

However, geothermal development could have certain negative impacts if appropriate mitigation actions and monitoring plants are not in place. Any large-scale construction and drilling operation will produce visual impacts on the landscape, create noise and wastes and affect local economies.

Environmental Impacts of Biomass

Biomass power plants share some similarities with fossil fuel power plants: both involve the combustion of a feedstock to generate electricity. Thus, biomass plants raise similar, but not identical, concerns about air emissions and water use as fossil fuel plants. Biomass power plants, like coal and natural gas-fired power plants, require water for cooling. Land use impacts from biomass power production are driven primarily by the type of feedstock: either a waste stream or an energy crop that is grown specifically for generating electricity.

There are global warming emissions associated with growing and harvesting biomass feedstock, transporting feedstock to the power plant, and burning or gasifying the feedstock. Transportation and combustion emissions are roughly equivalent for all types of biomass. It is understood that some biomass feedstock sources are associated with substantial global warming emissions. However, global warming emissions from the sourcing of biomass feedstock vary widely.

The energy transfers which occur in each type of energy resource.

Fossil fuels: chemical energy → thermal energy → kinetic energy → electrical energy

Solar energy: light & thermal energy → electrical energy

Nuclear: nuclear energy → thermal energy → kinetic energy → electrical energy

Wind: kinetic energy → electrical energy

Bio-fuel: chemical energy → thermal energy → mechanical energy → electrical energy

Hydroelectricity: gravitational potential energy → kinetic energy → electrical energy

Geothermal: thermal energy → kinetic energy → electrical energy

Tidal: kinetic energy → electrical energy

Reviewed Exercise

- What are consequences of development on the environment?

Key Words: global warming, air pollution, water pollution, greenhouse gas

SUMMARY

A **non-renewable resource** is a natural substance that is not replenished with the speed at which it is consumed.

Renewable energy sources (also called non-conventional energy sources) are continuously replenished by natural processes.

Solar energy is the radiant light and heat from the sun that has been harnessed by humans since ancient times using a range of ever-evolving technologies.

Conversion of light energy into electrical energy is based on a phenomenon called **photovoltaic effect**.

Wind energy is the energy generated or produced by harnessing the power of the wind.

The flowing water and water stored in huge dams is also a very important source of energy which is known as **hydroelectric energy**.

Geothermal energy is a type of renewable energy taken from the Earth's core (the internal heat of the earth) and it is one of the oldest types of natural sources of heat.

Fossil fuels, such as coal, oil and natural gas, are important non-renewable sources of energy.

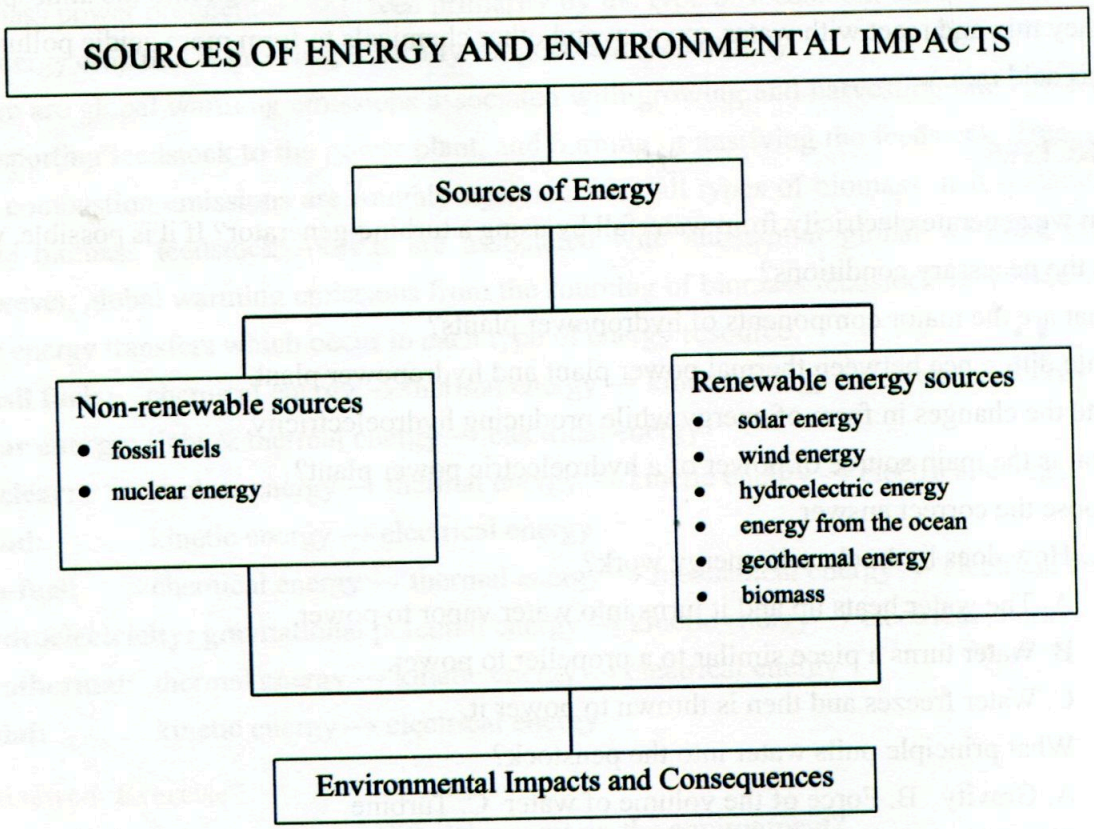
Acid rain is caused by a chemical reaction that begins when compounds like sulfur dioxide and nitrogen oxides are released into the air. These substances can rise very high into the atmosphere, where they mix and react with water, oxygen, and other chemicals to form more acidic pollutants, known as acid rain.

EXERCISES

1. Can we generate electricity from waterfall by using a turbine-generator? If it is possible, what are the necessary conditions?
2. What are the major components of hydropower plants?
3. Write difference between thermal power plant and hydropower plant.
4. State the changes in form of energy while producing hydroelectricity.
5. What is the main source of power of a hydroelectric power plant?
6. Choose the correct answer.
   (i) How does hydroelectric energy work?
   A. The water heats up and it turns into water vapor to power.
   B. Water turns a piece similar to a propeller to power.
   C. Water freezes and then is thrown to power it.
   (ii) What principle pulls water into the penstock?
   A. Gravity B. Force of the volume of water C. Turbine
   (iii) Which is true for hydroelectric energy?
   A. It does not costs to build. B. Water everywhere, it can be used everywhere.
   C. It rains all the time
7. Why is geothermal energy considered as a renewable resource?
8. How is geothermal electricity produced?

9. What are the advantages of geothermal energy compared to other energies?
10. An important hydroelectric plant has a head of 100 m and water volumetric flow rate of 10 000 $m^3 s^{-1}$. The turbine generator efficiency is 60 %. What is the maximum power that the plant can produce? (Density of water = 1000 kg m^{-3})
11. A dam which maximum head of 200 m produces to 2000 MW electrical power. What is the rate of falling water on the turbines? The turbine generator efficiency is 50 %. (Density of water = 1000 kg m^{-3})
12. The wind is blowing at 10 m s^{-1}, how much total power produce by the wind turbine if the blades are 45 m long? Density of air is 1.3 kg m^{-3}.
13. Why is it better to use more renewable energy resources rather than non-renewable resources?

CONCEPT MAP



CHAPTER 5

HEAT TRANSFER AND THERMODYNAMICS

Heat is the form of energy that can be transferred from one system to another as a result of temperature difference. In the previous levels, we have already learnt the concepts of heat, temperature, law of heat exchange and phase change. In this chapter, we are going to study heat transfer processes and the basis of thermodynamics. Heat transfer and thermodynamics are two complementary branches of science. Heat transfer is a discipline of thermal engineering that concerns the generation, use, conversion, and exchange of thermal energy between physical systems. Thermodynamics deals with equilibrium states by outlining the relationship between energy and work.

Learning Outcomes

It is expected that students will

- explain some phenomena that involve conductive, convective, and radiative heat transfer.
- understand thermal conductivity and emissivity, total emissive power and blackbody.
- solve problems using the formulas for conduction and radiation.
- solve problems on the relationships between heat transfer, time, and rate of heat transfer.
- know the difference between internal energy and heat transfer.
- realize some uses of properties of thermodynamics.
- examine heat engines and their efficiency.

5.1 HEAT TRANSFER

One of the basic requirement for heat transfer is the presence of a temperature difference. There can be no net heat transfer between two systems that are at the same temperature. There are three basic ways of heat transfer; namely, heat conduction, heat convection and heat radiation. In fluids, heat is mainly transferred by convection, in which the motion of the fluid itself carries heat from one place to another. In solid, heat is transferred by conduction process in which heat energy is transferred among the adjacent parts of the material without the motion of the material itself. The third way to transfer heat energy is by radiation process in the form of electromagnetic waves. The three basic heat transfer processes are illustrated in Figure 5.1.

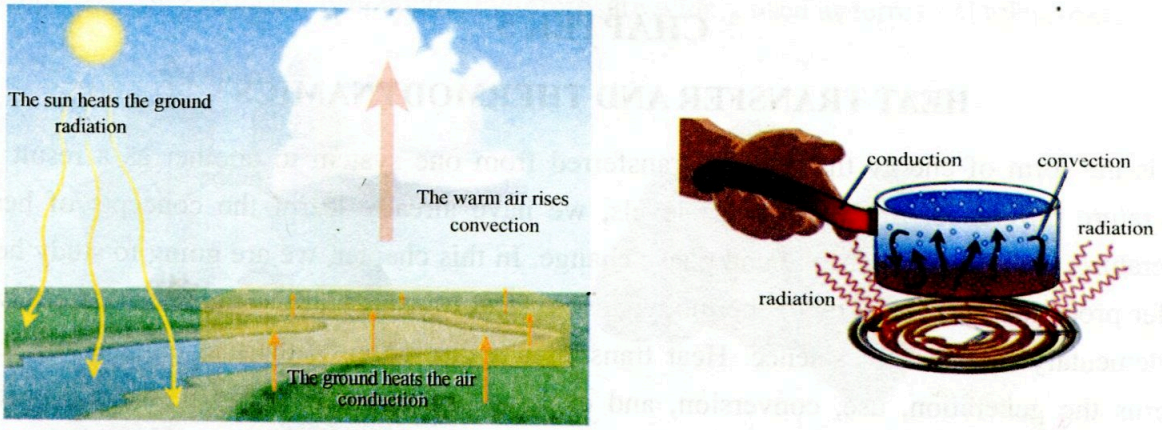


Figure 5.1 Three heat transfer processes

Heat Conduction

Heat conduction is one mode of energy transfer. The individual parts of a medium do not move as a whole in heat conduction. Heat energy is transferred by successive distribution among the adjacent parts in heat conduction.

When heat is transferred via conduction, heat energy is transferred internally by vibrations of atoms and molecules. Metals have many free electrons. Since these free electrons can carry heat, metals are generally very good conductors of heat. For example, if a tea spoon is put into a hot cup of tea, the spoon handle becomes hot. At first the end of the spoon placed in the hot tea gains heat energy. Then the handle end of the spoon becomes hot by successive distribution of heat energy among the adjacent parts. However, the spoon which acts as a medium for heat transfer, does not move at all.

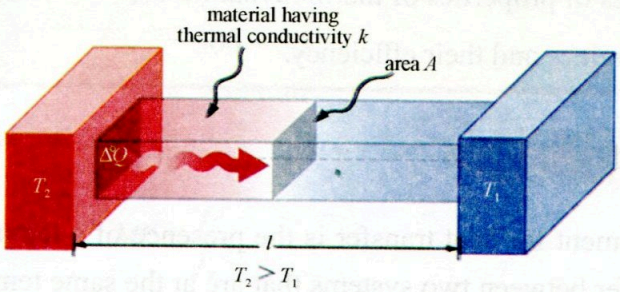


Figure 5.2 Heat conduction process

The rate of heat conduction through a material of cross-sectional area A is proportional to the temperature difference ΔT, but is inversely proportional to the thickness l of the material as shown in Figure 5.2. The heat transfer or conduction rate H_{cond} is

$$H_{cond} = k A \frac{T_2 - T_1}{l} \qquad (5.1)$$

where k is the thermal conductivity of the material, which is a measure of the ability of a material to conduct heat.

In SI units, k is watt per metre per kelvin (W m^{-1} K^{-1}). The ratio $\frac{T_2 - T_1}{l}$ is called temperature gradient. It is also expressed in differential form as $\frac{dT}{dx}$.

In one dimensional form,

$$\frac{H}{A} \propto -\frac{dT}{dx}$$

$$\frac{H}{A} = -k\frac{dT}{dx}$$

where $\frac{H}{A}$ is heat flow per unit area and H is heat current.

This mathematical model of heat conduction was first formulated by the French physicist Jean Baptiste Joseph Fourier in his Fourier's law of heat conduction. This law states that the rate of heat flow per unit area (heat flux density) is proportional to the magnitude of temperature gradient. Negative sign indicates that temperature T must decrease as distance x increases.

In the SI units, heat flux density is watt per metre squared (W m^{-2}) and heat current is watt (W).

The thermal conductivity of a material can be defined as the rate of heat transfer per unit area per unit temperature gradient. A high value for thermal conductivity indicates that the material is a good thermal conductor, and a low value indicates that the material is a poor thermal conductor or thermal insulator. The thermal conductivities of some common materials at room temperature are given in Table 5.1.

Table 5.1 The thermal conductivities of some materials at room temperature

Materials	Thermal conductivity, k (W m^{-1} K^{-1})	Materials	Thermal conductivity, k (W m^{-1} K^{-1})
Silver	429	Water	0.613
Copper	401	Human skin	0.37
Gold	317	Wood (oak)	0.17
Aluminum	237	Helium (gas)	0.152
Iron	80.2	Soft rubber	0.13
Mercury	8.54	Glass fiber	0.043
Glass	0.78	Air (gas)	0.026
Brick	0.72	Urethane, rigid foam	0.026

Note that materials such as copper and silver that are good electric conductors are also good heat conductors, and have high values of thermal conductivity. Materials such as rubber, wood, and styrofoam are poor conductors of heat and have low conductivity values.

Example 5.1 A person walking at a regular speed generates heat at the rate of 0.07 W. If the surface area of the body is 1.5 m^2 and heat is to be generated 0.03 m below the skin, what should be the temperature difference between the skin and interior of the body if the heat is to be conducted to the surface of the skin? ($k = 5 \times 10^{-5}$ W m^{-1} K^{-1})

Assume that heat is conducted through a small tissue.

$H_{cond} = 0.07$ W, $A = 1.5$ m^2, $l = 0.03$ m, $k = 5 \times 10^{-5}$ W m^{-1} K^{-1}

$$H_{cond} = kA\frac{(T_2 - T_1)}{l}$$

Thus, $$(T_2 - T_1) = \frac{l\, H_{cond}}{kA}$$

$$= \frac{0.03 \times 0.07}{5 \times 10^{-5} \times 1.5} = 28\,\text{K} = 28\ ^{\circ}\text{C}$$

This value of temperature difference is not realistic in human beings. The temperature difference in body is only a few degrees. Actually, the major heat transfer in human beings is due to the flow of warm blood.

Heat Convection

Heat convection is the flow of heat through a fluid from places of higher temperature to places of lower temperature by movement of the fluid itself.

In heat convection, the transfer of energy occurs between an object and its environment due to fluid motion. In other words, the heat is carried by a flowing fluid from one place to another. Convection currents are set up in the fluid because the hotter part of the fluid is not as dense as the cooler part, so there is an upward buoyant force on the hotter fluid, making it rise while the cooler and denser fluid sinks as shown in Figure 5.3 (a). Although some heat may be transferred by conduction in liquids and gases, a much larger quantity of heat is carried by the motion of the fluid itself. In cold regions where rooms are heated by fire, heating is done by convection process. The fluid carrying the heat is the air in the room as shown in Figure 5.3 (b).

Convection occurs only in a liquid or a gas, never in a solid.

Buoyant force is what drives natural convection, where the pressure difference between air heated by heat sources and surrounding ambient air drives the hot air upwards and away from the heat source.

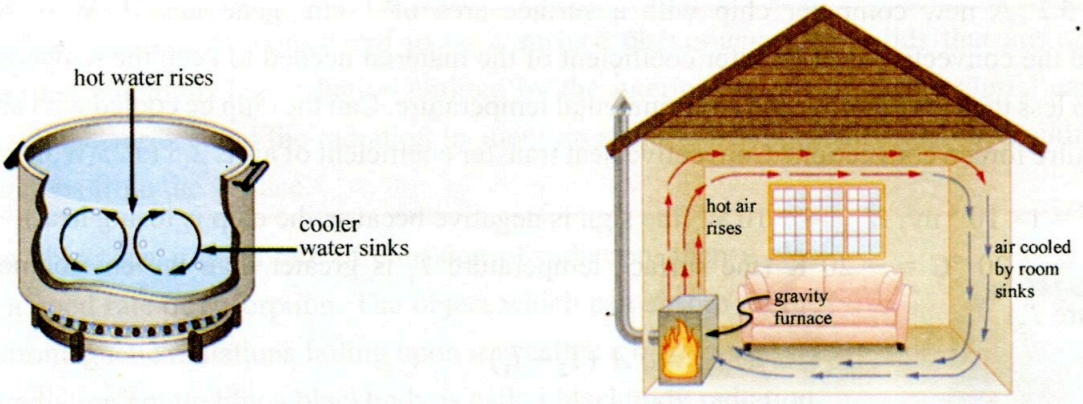


Figure 5.3 Heat convection process

The basic equation for the rate of convection heat transfer is known as Newton's law of cooling,

$$H_{conv} = q\,A\,(T_2 - T_1) \qquad (5.2)$$

where H_{conv} is the rate of heat convection, q is the convective heat transfer coefficient (heat convection constant), A is the surface area of the object being cooled or heated, T_2 is the temperature of the surrounding fluid, and T_1 is the surface temperature of the object. In SI units, q is expressed watt per metre squared per kelvin (W m^{-2} K^{-1}).

When $T_2 > T_1$, heat is gained by the object and H_{conv} is positive. When $T_2 < T_1$, heat is lost by the object and H_{conv} is negative.

Some of the weather conditions are created by heat convection as shown in Figure 5.4. One of the reasons why the weather is fair at the base of mountain ranges, at the sea coast, lakes and ponds is that the hot air in those regions rises and is replaced by cooler air which blows across the water surface. This process occurs due to heat convection. The heat loss by convection is important for living beings.

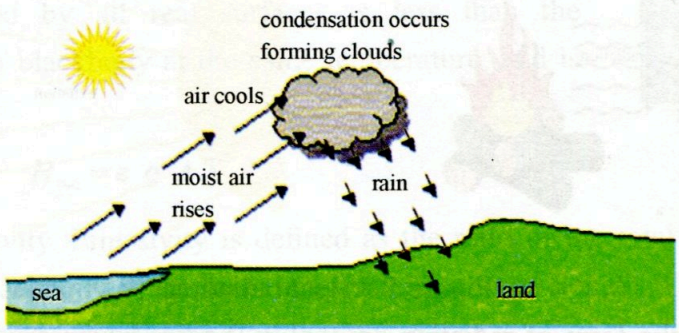


Figure 5.4 Creation of weather condition

The household hot water system is also designed based on the process of heat convection in liquid. In cold countries, many homes have a hot water tank. Hot water from the boiler rises and flows into the tank where it rises to the top. When you use a hot water tap at home, you draw hot water from the top of the tank.

Example 5.2 A new computer chip with a surface area of 1 cm^2 generates 10 W of heat. Determine the convective heat transfer coefficient of the material needed to keep the temperature of the chip less than 20 °C above the environmental temperature. Can the chip be cooled with air, or will it require forced convection? Convective heat transfer coefficient of air is 2.5 to 25 W m^{-2} K^{-1}.

A = 1 cm^2 = 1× 10^{-4} m^2, H_{conv} = − 10 W (the sign is negative because the chip is losing heat)
(T_2 − T_1) = − 20 °C = − 20 K (the surface temperature T_1 is greater than the environmental temperature T_2)

$$H_{conv} = q\,A\,(T_2 - T_1)$$

$$q = \frac{H_{conv}}{A\,(T_2 - T_1)}$$

$$= \frac{-10}{1\times10^{-4}\times(-20)} = 5000\ \text{W m}^{-2}\text{K}^{-1}$$

Since convective heat transfer coefficient of air is very much less than 5000 W m^{-2} K^{-1}, the chip cannot be cooled with air. Therefore it will require the forced convection.

Heat Radiation

Heat radiation is a transfer of heat that does not require a material medium. Heat is transferred by thermal radiation which are electromagnetic waves that pass through a medium or even vacuum. Unlike conduction and convection, the transfer of energy by radiation does not require an intervening medium. In fact, energy transfer by radiation is the fastest (at the speed of light) and it suffers no attenuation in a vacuum. This is how the energy of the sun reaches the earth by heat radiation as shown in Figure 5.5.

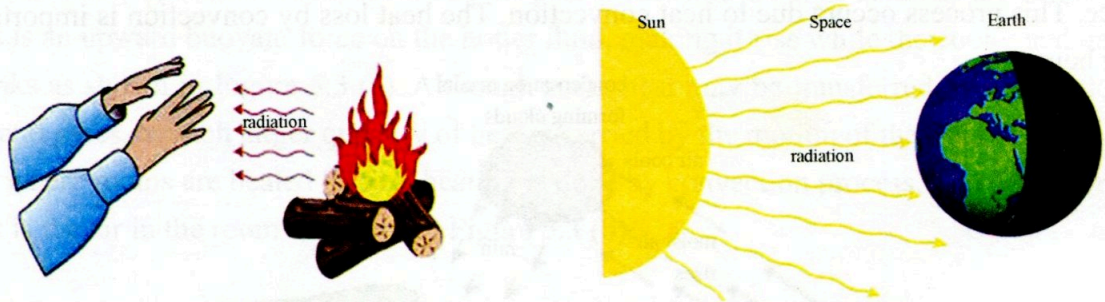


Figure 5.5 Heat radiation

Thermal radiation is the form of radiation emitted by bodies because of their temperature. All bodies at a temperature above absolute zero emit thermal radiation. All solids, liquids and gases emit or absorb radiation to varying degrees. Thermal radiation differs from other forms of electromagnetic radiation such as X-rays, gamma rays, microwave and radio wave that are not related to temperature.

Radiation is mostly considered to be a surface phenomenon for solids that are opaque to thermal radiation. The radiation emitted by the interior regions of such material can hardly reach the surface and the radiation incident on such bodies is usually absorbed within a few microns from the surface.

If an object has a good rate of emission of radiation, then it also has a good rate of absorption. The object which can absorb all the electromagnetic radiations falling upon it is called a blackbody and the radiation emitted by a blackbody is called blackbody radiation. The blackbody is not only a perfect absorber but also a best emitter of radiation.

Josef Stefan (1835- 1893)

The total emissive power is defined as the total radiant energy of different wavelengths emitted from unit area of a surface of a body in one second.

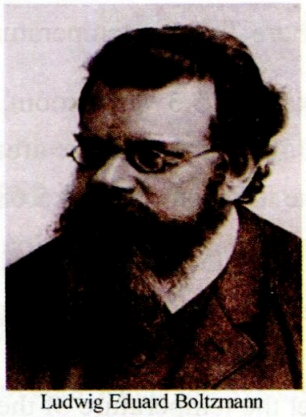
The total emissive power of a blackbody is directly proportional to the fourth power of absolute temperature.

$$\varepsilon_0 = \sigma T^4 \qquad (5.3)$$

This equation is known as Stefan-Boltzmann's law and σ is called Stefan's constant, $\sigma = 5.685 \times 10^{-8}$ W m^{-2} K^{-4}.

The maximum rate of radiation that can be emitted from a surface of area A (blackbody) at an absolute temperature T is given by the Stefan-Boltzmann's law.

$$H_{rad} = \sigma A T^4 \qquad (5.4)$$



Ludwig Eduard Boltzmann (1844 - 1906)

The radiation emitted by all real surfaces is less than the radiation emitted by a blackbody at the same temperature, and is expressed as

$$H_{rad} = e\ \sigma A T^4 \qquad (5.5)$$

where e is the emissivity. Emissivity is defined as the ratio of the total emissive power of an object to that of the blackbody, at the same temperature and wavelength. That is $e = \frac{\varepsilon}{\varepsilon_0}$. The total emissive power of objects other than a black body is $\varepsilon = e\varepsilon_0$.

Emissivity of an object depends to a large extent on how shiny it is. If an object reflects a lot of energy, it will absorb (and radiate) very little; if it reflects very little energy, it will absorb and radiate quite efficiently.

A black object is a good absorber and a good radiator whose emissivity is close to 1, whereas a white, clear, or silver coated object is a poor absorber and a poor radiator whose emissivity is much less than 1 as shown in Figure 5.6.

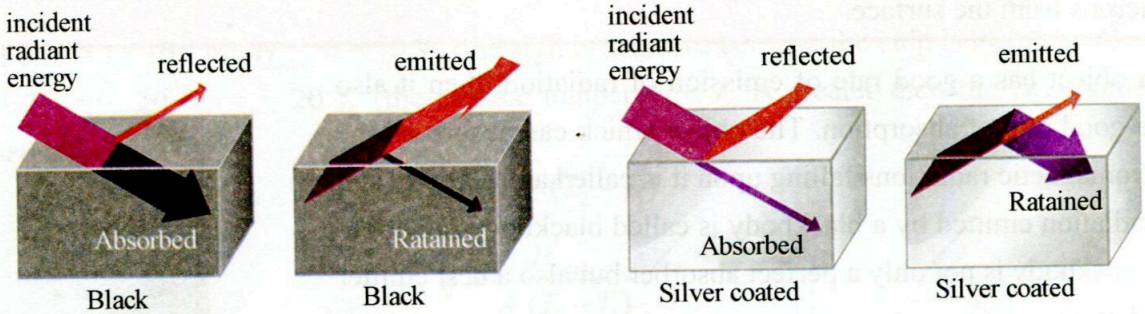


Figure 5.6 Good absorber, good radiator and poor absorber, poor radiator

Note that the emissivity of an object depends on the wavelength of radiation. A shiny object may reflect a great deal of visible light, but it may be a good absorber (emitter) of radiation of a different wavelength, such as ultraviolet or infrared light.

The net energy change is simply the difference between the radiated energy and the absorbed energy. This can be expressed as the net radiation per unit area of the surface per second is,

$$\frac{H_{net}}{A} = e\sigma(T_2^4 - T_1^4)$$

where T_2 is the temperature of the object and T_1 is the temperature of the surroundings.

Example 5.3 In a room, the temperature of water in a kettle is 80 °C . The room temperature is 30 °C and the surface area of the kettle is 0.02 m^2. If the emissivity is 0.9 find the rate of heat loss due to radiation. ($\sigma = 5.685 \times 10^{-8}$ W m^{-2} K^{-4})

The kettle is emitting heat itself and at the same time absorbing heat radiated from the surrounding. It is emitting heat at its temperature 80 °C and absorbing heat from the surrounding at the temperature 30 °C.

Let the temperature of the kettle be T_2 and the surrounding's temperature is T_1.

$T_2 = 80 + 273 = 353$ K, $T_1 = 30 + 273 = 303$ K

$A = 0.02$ m^2, $\sigma = 5.685 \times 10^{-8}$ W m^{-2} K^{-4}, e = 0.9,

The rate of heat radiation $H_{rad} = e\,\sigma\,A\,T^4$

The rate of heat radiated from the kettle at the temperature T_2 is

$$H_{out} = e\sigma A T_2^4$$
$$= 0.9 \times 5.685 \times 10^{-8} \times 0.02 \times (353)^4$$
$$= 15.89 \text{ W}$$

The rate of heat absorbed by the kettle at the temperature T_1 is

$$H_{in} = e\sigma AT_1^4$$
$$= 0.9 \times 5.685 \times 10^{-8} \times 0.02 \times (303)^4$$
$$= 8.63 \text{ W}$$

Thus, the net rate of heat lost by the kettle is

$$H_{lost} = H_{out} - H_{in}$$
$$= 15.89 - 8.63$$
$$= 7.26 \text{ W}$$

Reviewed Exercise

1. Some electric stoves have a flat ceramic surface with heating elements hidden beneath. A metallic pot placed directly over a heating element will be heated more effectively. Why is ceramic, with a conductivity less than that of a metal but greater than that of a good insulator, an ideal choice for the stove top?
2. Why is heat transfer take place between an object and its surrounding?
3. What is the relation between the total emissive power of an object and total emissive power of black body when both are at same temperature?

Key Words: heat conduction, heat convection, radiation, thermal conductivity, emissivity, blackbody

5.2 THERMODYNAMICS

Thermodynamics is the study of systems involving energy in the form of heat and work. A good example of a thermodynamic system is gas confined by a piston in a cylinder. If the gas is heated, it will expand, doing work on the piston; this is one example of how a thermodynamic system can do work.

Thermal equilibrium is an important concept in thermodynamics. When two systems are in thermal equilibrium, there is no net heat transfer between them. This occurs when the systems are at the same temperature. In other words, systems at the same temperature will be in thermal equilibrium with each other.

The Zeroth Law of Thermodynamics

The zeroth law of thermodynamics is a simple definition of thermodynamic equilibrium. It states that;

When two objects are separately in thermodynamic equilibrium with a third object, they all are in thermodynamic equilibrium with each other.

First Law of Thermodynamics

The first law of thermodynamics is an application of the conservation energy for heat and thermodynamic processes. The first law of thermodynamics relates the change in internal energy of a system to heat added and the work done by the system.

The change in internal energy ΔU of a system is equal to the net heat flow Q into the system minus the total work W done by the system.

In symbols,
$$\Delta U = Q - W \tag{5.6}$$

Q is positive if heat is added to the system, and negative if heat is removed. W is positive if work is done by the system, and negative if work is done on the system.

Internal Energy

The intenal energy U of a system is defined as the sum of the energies of all the constituent particles. The internal energy of ideal gases depends only on temperature, not on volume or pressure. The internal energy of an ideal monoatomic gas was given as;

$$U = \frac{3}{2} N k_B T = \frac{3}{2} N \frac{RT}{N_A}$$

$$U = \frac{3}{2} n R T \tag{5.7}$$

where N = number of particles of an ideal gas, N_A = Avogadro's number, T = absolute temperature, R = universal (ideal) gas constant, $n = \frac{N}{N_A}$ = number of moles, and $k_B = \frac{R}{N_A}$ = Boltzmann constant.

Note that internal energy of an ideal gas is directly proportional to the absolute temperature.

Heat transfer and work are both energy in transit. Neither is stored as such in a system. However, both can change the internal energy U of a system. Internal energy is a form of energy completely different from either heat or work.

Thermodynamics Processes

A thermodynamic process may be defined as a change from one equilibrium state of a system to another. During such a process, a system starts from an initial state i (pressure p_i, volume V_i and temperature T_i) passes through various quasistatic states to a final state f (pressure p_f, volume V_f and temperature T_f).

In such a process energy may be transferred from (or) into the system and also work can be done by (or) on the system. There are various thermodynamic processes that are of importance in engineering of heat engines.

In thermodynamics, a quasi-static process (also known as a quasi-equilibrium process) is a thermodynamic process that happens slowly enough for the system to remain in internal physical thermodynamic equilibrium.

1. Isobaric Process

An isobaric process occurs at constant pressure. An example of an isobaric process is a gas, being slowly heated or cooled, confined by a piston in a cylinder. The heat Q, the work W, or the change in internal energy ΔU are nonzero in an isobaric process. The work done by the system in an isobaric process is simply the pressure multiplied by the change in volume. In an isobaric process the pressure is constant and the work is equal to the area of the shaded rectangle as shown in Figure 5.7 (pressure Vs volume graph).

In an isobaric process, the work is, $$W = p\int_{V_i}^{V_f} dV = p\ (V_f - V_i) \qquad (5.8)$$

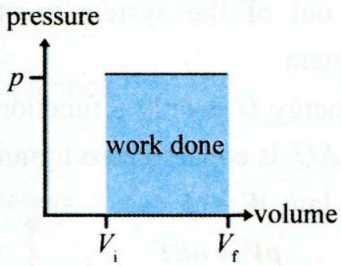


Figure 5.7 Isobaric process

2. Isochoric Process

An isochoric process occurs at constant volume. An example of an isochoric (isovolumetric) process is heating of a gas in a rigid, closed container. When the volume of a thermodynamic system is constant, no work is done by the system because $W = p\,\Delta V = 0$ as shown in Figure 5.8. According to the first law, the change in internal energy must equal any heat flowing into or out of the system. That is, $\Delta U = Q$.

Since $Q = m\,c_v\,\Delta T$, $\Delta U = m\,c_v\,\Delta T$ where c_v is specific heat capacity at constant volume.

However, $c_v = \dfrac{C_v}{M}$, where C_v is molar heat capacity at constant volume and M is molar mass of the gas $\left(n = \dfrac{m}{M}\right)$, n is number of moles.

The change in internal energy is expressed in terms of molar heat capacity C_v,

$$\Delta U = Q = n\ C_v\ \Delta T \qquad (5.9)$$

For an ideal monatomic gas

$$\Delta U = \frac{3}{2} n\ R\,\Delta T \qquad (5.10)$$

where $C_v = \dfrac{3}{2}R$.

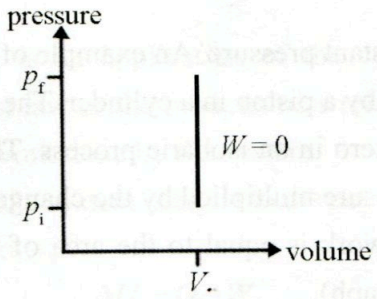


Figure 5.8 Isochoric process

3. Isothermal Processes

An isothermal process occurs at a constant temperature (i.e. $\Delta T = 0$). For a process to be isothermal, any heat flow into or out of the system must occur slowly to maintain the thermodynamic equilibrium of the system.

However, for an ideal gas, internal energy U is only a function of the temperature. In this case, internal energy is constant; therefore, ΔU is equal to zero for an isothermal process.

Since $\Delta U = 0$, according to the first law, $W = Q$.

Using the ideal gas law $pV = nRT$

The work, represented by the area under curve as shown in Figure 5.9, can be calculated as follow.

$$W = \int_{V_i}^{V_f} p\, dV = \int_{V_i}^{V_f} \frac{n\,R\,T}{V} dV = n\,R\,T \int_{V_i}^{V_f} \frac{1}{V} dV$$

$$W = n\,R\,T\,\ln\left(\frac{V_f}{V_i}\right) \qquad (5.11)$$

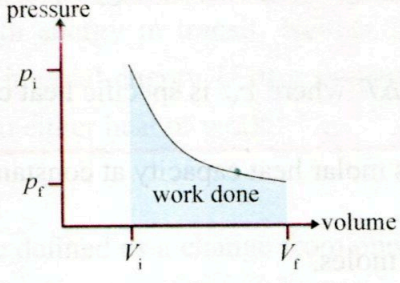


Figure 5.9 Isothermal process

If the system does the work, the energy comes from heat flowing into the system from the reservoir; if the work is done on the system, heat flows out of the system to the reservoir.

4. Adiabatic Processes

In an adiabatic process, there is no heat transfer into or out of the system, i.e. $Q = 0$. This process can be done either by enclosing the system in a thermal insulator (adiabatic container) (or) by carrying out the process so quickly that there is no enough time for heat transfer. [Figure 5.10]

The first law of thermodynamics $\Delta U = Q - W$ is thus reduced to $\Delta U = -W$.

For an ideal gas, $\Delta U = \frac{3}{2} n\ R\,\Delta T$.

Therefore, work done $W = -\Delta U = -\frac{3}{2} n\ R\,\Delta T$

$$W = \frac{3}{2} n\,R\,(T_i - T_f) \qquad (5.12)$$

Using the ideal gas law in the form $pV^{\gamma} = K$ (constant), the wok done W can also be expressed in terms of pressure and volume as,

$$W = \frac{1}{\gamma - 1}(p_i V_i - p_f V_f) \qquad (5.13)$$

where γ is the ratio of the heat capacities, $\gamma = \frac{C_P}{C_V}$.

C_p is molar heat capacity at constant pressure.

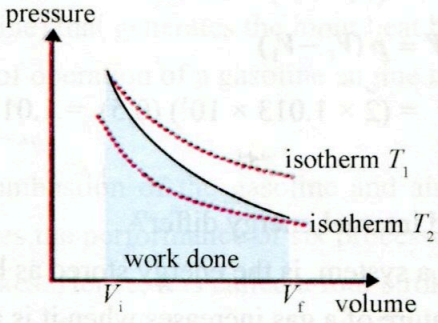


Figure 5.10 Adiabatic process compare with isothermal process

Example 5.4 Find the change in internal energy of one mole of an ideal gas when its temperature changes from 0 °C to 100 °C.

$\Delta T = (100 - 0) = 100$ °C $= 100$ K, $n = 1$ mol, $R = 8.3143$ J mol^{-1} K^{-1}

The change in internal energy,

$$\Delta U = \frac{3}{2} n\,R\,\Delta T$$

$$= \frac{3}{2} \times 1 \times 8.3143 \times 100 = 1.25 \times 10^{3}\ \text{J}$$

Example 5.5 One gram of water (1cm^3) becomes 1671 cm^3 of steam when boiled at a pressure of 1 atm. The heat of vaporization at this pressure is 2265 J g^{-1}. Compute the work done and the increase in internal energy.

$m = 1$ g , $p = 1$ atm $= 1.013 \times 10^5$ Pa, $L_v = 2265$ J g^{-1}

$(V_f - V_i) = 1671 - 1 = 1670$ cm^3 $= 1670 \times 10^{-6}$ m^3

For isobaric process;

$$W = p(V_f - V_i)$$

$$= 1.013 \times 10^5 \times 1670 \times 10^{-6} = 169 \text{ J}$$

Heat required one gram of water to vaporize;

$$Q = m L_v$$

$$= 1 \times 2265 = 2265 \text{ J}$$

According to first law of thermodynamics

$$\Delta U = Q - W$$

$$= 2265 - 169 = 2096 \text{ J}$$

Example 5.6 In an isobaric process, a volume of a gas is expressed by absorption of heat at constant pressure 2 atm. If the increment in volume is 0.5 m^3, then find the work done by the system.

$p = 2 \text{ atm} = 2 \times 1.013 \times 10^5 \text{ Pa}$

$$\Delta V = (V_f - V_i) = 0.5 \text{ m}^3$$

In an isobaric process, $W = p(V_f - V_i)$

$$= (2 \times 1.013 \times 10^5)(0.5) = 1.013 \times 10^5 \text{ J}$$

Reviewed Exercise

1. How do heat transfer and internal energy differ?
2. When heat transfers into a system, is the energy stored as heat? Explain briefly.
3. Explain why the temperature of a gas increases when it is adiabatically compressed.

Key Words: thermodynamics, internal energy, isobaric process, isothermal process, isochoric process (isometric process), adiabatic processes

5.3 HEAT ENGINES

Heat engine is a system that converts heat to mechanical energy, which can then be used to do mechanical work. It operates continuously and only heat and work may pass across its boundaries. A series of processes in which a system is brought back to its initial state is called a cycle. The purpose of a heat engine is to deliver work continuously to the outside by performing the same cycle over and over again as illustrated in Figure 5.11.

Thermal efficiency for a heat engine is the ratio of work output to heat input.

$$\text{Thermal efficiency} = \frac{\text{work output}}{\text{heat input}}$$

$$\eta = \frac{W}{Q_H} \qquad (5.14)$$

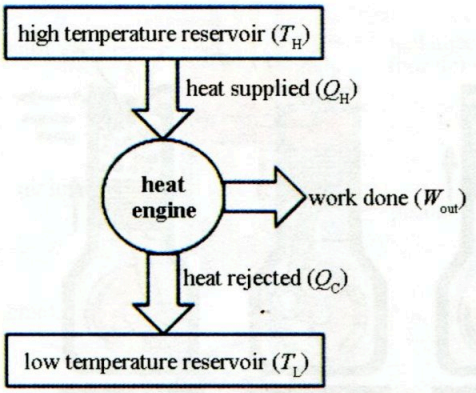


Figure 5.11 Thermodynamic cycle of heat engine

Applying the first law of thermodynamic to one complete cycle, remembering that there is no net change of internal energy ($\Delta U = 0$), we get, $Q_H - Q_C = W$.

The transformation of heat into work is usually accomplished in practice by two general types of engine; the internal combustion engine and the steam engine. In both engines, a gas or a mixture of gases contained in a cylinder undergoes a cycle. There are two types of internal combustion engine (gasoline engine and diesel engine) that generates the input heat by the combustion of fuel within the engine itself. The principle of operation of a gasoline engine is described as below.

1. Gasoline Engines

In the gasoline engines, the combustion of the gasoline and air takes place explosively by an electric spark. The cycle involves the performance of six processes, of which four require motion of the piston and are called strokes. Hence, it is called a four stroke engine.

(i) **Intake stroke:** A mixture of gasoline vapour and air is drawn into the cylinder by the suction stroke of the piston.

(ii) **Compression stroke:** The mixture of gasoline vapour and air is compressed until its pressure and temperature rise considerably. This is accomplished by the compression of the piston.

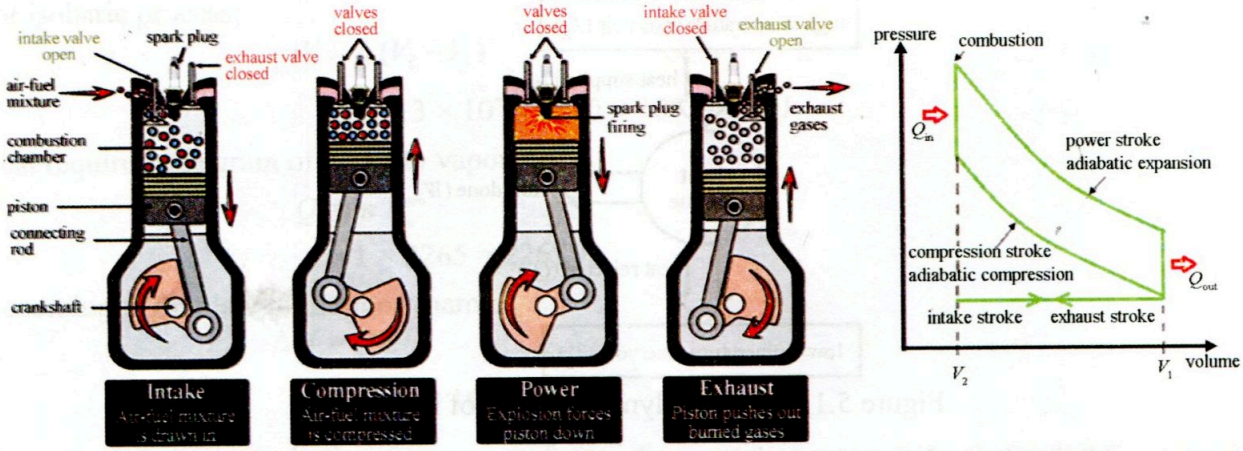
(iii) **Explosion:** Combustion of the hot mixture is caused to take place very rapidly by an electric spark. The combustion products attain a very high pressure and temperature, but the volume remains unchanged. The piston does not move during this process.

(iv) **Power stroke:** The hot combustion products expand and push the piston out, thus suffering a drop in pressure and temperature.

(v) **Valve exhaust:** The combustion products at the end of the power stroke are still at a higher pressure and temperature than the outside. An exhaust valve allows some gas to escape until the pressure drops to that of the atmosphere. The piston does not move during this process.

(vi) **Exhaust stroke:** The piston pushes almost all the remaining combustion products out of the cylinder by exerting a pressure sufficiently larger than that of the outside pressure.

The four strokes cycle and the *p-V* diagram of a gasoline engine is shown Figure 5.12 (a) and (b).



(a) (b)

Figure 5.12 (a) The four strokes cycle and (b) *p-V* diagram of a gasoline engine

Theoretical efficiency of a gasoline engine is

$$\eta = 1 - \frac{1}{\left(\frac{V_1}{V_2}\right)^{\gamma-1}} \qquad (5.15)$$

where $\frac{V_1}{V_2}$ = compression ratio, and $\gamma = \frac{C_p}{C_v}$ = ratio of molar heat capacities. For a typical compression ratio of 8 and with $\gamma = 1.4$, $\eta = 56\%$ for an ideal gasoline engine.

2. Diesel Engine

The diesel engine operates on either a two-stroke or four-stroke cycle; however, unlike the spark-ignition gasoline engine, the diesel engine induces only air into the combustion chamber on its intake stroke. The operation of a two-stroke diesel engine is explained below.

When the piston is at the top of its travel, the cylinder contains a highly compressed air. Diesel fuel is sprayed into the cylinder by the injector (nozzle) and immediately ignites because of the heat and pressure inside the cylinder as shown in Figure 5.13.

The pressure created by the combustion of the fuel drives the piston downward. This is the power stroke. As the piston nears the bottom of its stroke, all of the exhaust valves open. Exhaust gases rush out of the cylinder, relieving the pressure.

As the piston is at the bottom, it uncovers the air intake ports. Pressurized air fills the cylinder, forcing out the remainder of the exhaust gases. The exhaust valves close and the piston starts traveling back upward, re-covering the intake ports and compressing the fresh air. This is the compression stroke. As the piston nears the top of the cylinder, the cycle repeats.

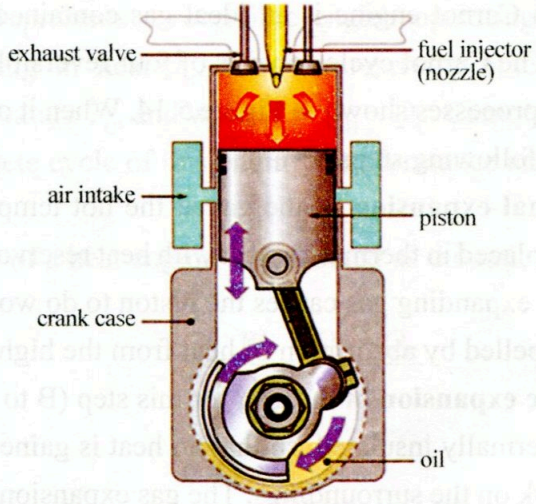


Figure 5.13 The two strokes cycle of diesel engine

Steam engine is operating between the steam temperature 373 K and outside room temperature 300 K. Gasoline engine is operating between petrol burning temperature 680 K and outside room temperature 300 K. Diesel engine is operating between the diesel burning temperature 873 K and outside room temperature.

3. Carnot Heat Engine

In 1824, Nicolas Leonard Sadi Carnot, French engineer, described a hypothetical engine known as Carnot engine. It is a heat engine operates on an ideal reversible cycle (Carnot cycle) between two heat reservoirs.

Carnot engine sets an upper limit on the efficiency of all other real engines, such as steam engines, gasoline and diesel engines. All real engines are less efficient than Carnot engine because they do not operate through a reversible cycle, and efficiency is further reduced due to friction and energy losses by conduction.

Nicolas Leonard Sadi Carnot
(1796 -1832)

In a reversible process, the system undergoing the process can be returned to its initial condition along the same path on the p-V diagram, and every point along this path is in an equilibrium state. A process that does not satisfy these requirements is an irreversible process.

The most important condition for a process to be reversible in thermodynamics is that the process should be carried out infinitesimally slowly and the driving force should be small enough to maintain a state of equilibrium.

The working substance of a Carnot engine is an ideal gas contained in a cylinder fitted with a movable piston at one end. The Carnot cycle consists of four reversible processes; two isothermal processes and two adiabatic processes shown in Figure 5.14. When it operates as a heat engine, the Carnot cycle consists of the following steps.

Step 1: Reversible isothermal expansion of the gas at the hot temperature T_H (isothermal heat addition). The gas is placed in thermal contact with heat reservoir at temperature T_H. During this step (A to B) the expanding gas causes the piston to do work on the surroundings. The gas expansion is propelled by absorption of heat from the high temperature reservoir.

Step 2: Reversible adiabatic expansion of the gas, for this step (B to C). The piston and cylinder are assumed to be thermally insulated, so that no heat is gained or lost. The gas continues to expand, doing work on the surroundings. The gas expansion causes it to cool to the cold temperature T_C.

Step 3: Reversible isothermal compression of the gas at the cold temperature T_C. (isothermal heat rejection) (C to D). The gas is in thermal contact with low reservoir at temperature T_C. Now the surroundings do work on the gas, compressing it and causing heat to flow out of the gas to the low temperature reservoir.

Step 4: Reversible adiabatic compression of the gas (D to A), once again the piston and cylinder are assumed to be thermally insulated. During this step, the surroundings do work on the gas, compressing it and causing the temperature to rise to T_H. At this point the gas is in the same state as at the start of step 1.

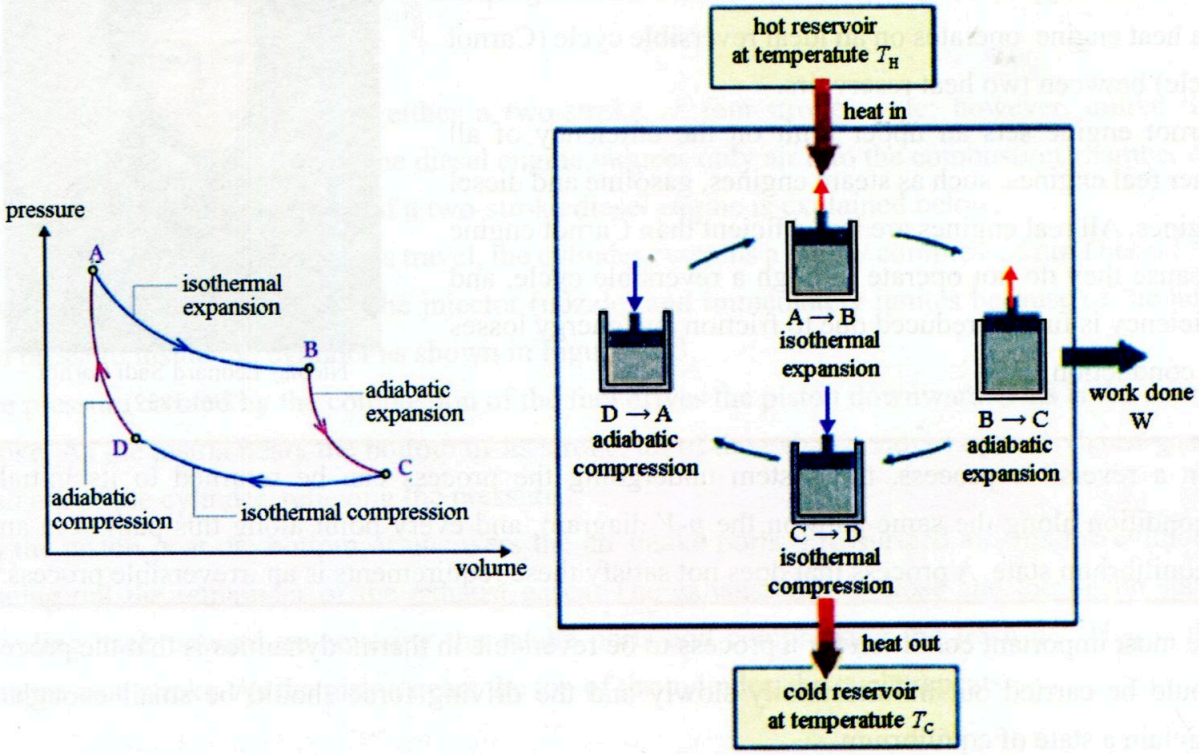


Figure 5.14 The four reversible processes of Carnot cycle

In Figure 5.14, the circle represents a reversible engine operating between two reservoirs at temperatures T_H and T_C, taking in heat Q_H, from the reservoir at temperature T_H, rejecting heat Q_C to the reservoir at temperature T_C, and work $W = Q_H - Q_C$. Carnot analyzed the transformation of energy during one complete cycle of this engine's performance and determined the conditions for maximum efficiency.

Since the thermal efficiency of a heat engine η is defined as the ratio of the work output W to the heat input Q_H.

$$\eta = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H}$$

$$\eta = 1 - \frac{Q_C}{Q_H} \tag{5.16}$$

Kelvin proposed that we use a Carnot engine as a thermometer and define the ratio of two temperatures as the the ratio of the heat absorbed by the engine to the heat rejected, when the engine is operated between reservoirs at these temperatures.

$$\frac{Q_C}{Q_H} = \frac{T_C}{T_H}$$

$$\eta_C = 1 - \frac{T_C}{T_H} \tag{5.17}$$

Although Carnot engine is not 100% efficient, but it is more efficient than other machine in transforming heat into work.

It should be noted that the temperatures of the reservoirs in this case are absolute temperature.

Carnot's theorem

Carnot's theorem states that no engine operating between two heat reservoirs can be more efficient than a Carnot engine operating between the same reservoirs.

$$\eta_C = 1 - \frac{T_C}{T_H}$$

Thus, this equation gives the maximum efficiency possible for any engine using the corresponding temperatures. A corollary to Carnot's theorem states that: all reversible engines operating between the same heat reservoirs are equally efficient.

Example 5.7 An engine takes in 9220 J and does 1750 J of work each cycle while operating between 689 °C and 397 °C. (a) What is its actual efficiency? (b) What is its maximum theoretical efficiency?

$Q_H = 9220$ J, $W = 1750$ J, $T_H = 689$ °C + 273 = 962 K, $T_C = 397$ °C + 273 = 670 K

Actual efficiency $\eta = \frac{W}{Q_H} = \frac{1750}{9220} = 0.19 = 19\ \%$

Maximum efficiency $\eta_C = 1 - \frac{T_C}{T_H}$

$$= 1 - \frac{670}{962} = 0.304 = 30.4\ \%$$

Reviewed Exercise

- Is a temperature difference necessary to operate a heat engine? Explain why or why not.

Key Words: heat engine, diesel engine, gasoline engine, Carnot engine, Carnot cycle, stroke

SUMMARY

The rate of heat flow (**heat current**) H is the amount of heat energy transferred per unit time.

Heat conduction is one mode of energy transfer. The individual parts of a medium do not move as a whole in heat conduction.

Heat convection is the flow of heat through a fluid by movement of the fluid itself.

Heat radiation is a transfer of heat that does not require a material medium. Heat is transferred by thermal radiation which are electromagnetic waves that pass through medium or even vacuum. The object which can absorb all the electromagnetic radiations falling upon it is called a **blackbody** and the radiation emitted by a blackbody is called **blackbody radiation**.

Heat engine is a system that converts heat to mechanical energy, which can then be used to do mechanical work.

Thermal efficiency of a heat engine η is defined as the ratio of the work output W to the heat input Q_H.

EXERCISES

1. What is thermal conductivity? Express its unit in SI system.
2. One end of a poker is placed in fire. After some time, the other end becomes hot. Explain how heat is transferred along the poker. Name the method of heat transfer in this case.
3. If a person wearing ordinary clothes travels out into space, the liquid in the body will boil. Why? Explain how a space suit can prevent this effect.
4. In cold regions it is seen that birds on the branches of trees often ruffle their feathers. Explain the reason why the birds feel warm by ruffling their feathers.
5. How does a blanket wrapped round your body keep you warm on a cold day?

6. Explain with a diagram why an air conditioner should be best positioned high, near the ceiling of a room.
7. Why do we generally wear dark clothes in the winter and light-colored clothes in the summer?
8. Can you warm a kitchen in winter by leaving the oven door open? Can you cool the kitchen on a hot summer day by leaving the refrigerator door open? Explain.
9. The area and thickness of a glass plate of a window are 0.25 m^2 and 4 mm respectively. The temperature of inside surface of glass plate is 25 °C and its outside temperature is 26 °C. Find the amount of heat that passes through the glass plate in one hour. The thermal conductivity of glass is 0.78 W m^{-1} K^{-1}.
10. How much heat per second is conducted through a wooden wall of area 25 m^2 and thickness 0.04 m if the temperature inside is 20 °C and the temperature outside is –10 °C ? The thermal conductivity of wood is 8.37×10^{-2} W m^{-1} K^{-1}.
11. The filament of a 100 W electric bulb is made of tungsten. The emissivity of tungsten is 0.3 and its length is 0.2 m. Find the diameter of the filament, if its temperature is 3000 K when the bulb is switched on. ($\sigma = 5.685 \times 10^{-8}$ W m^{-2} K^{-4})
12. The temperature of the filament is 2500 K when the bulb is switched on. The diameter of the filament is 0.1 mm and it is made of metal of emissivity 0.35. If the power is 40 W find the length of the filament.
13. From calculations based on the radiation measurement of solar energy falling on the earth it is found that sun is radiating energy at a rate of 62.5 MW m^{-2}. Assuming that the sun is emitting energy as a blackbody, find the temperature of the surface of the sun.
14. If the rate of energy radiation from a black body of area 100 cm^2 is 42 W, find the temperature of that blackbody.
15. Compare the rates of energy radiation of a blackbody at temperatures 327 °C and at 27 °C.
16. Mention the most important condition for a process to be reversible in thermodynamics.
17. For a gas system, how much work is done in an isometric process? Explain.
18. An insulated system takes in 6.5 kcal of heat and has 2000 J of work done on it. What is the change in internal energy of the system? (1 kcal = 4184 J)
19. A heat engine undergoes a process in which its internal energy increases by 275 J while it is doing 360 J of work. How much heat is taken in (or given out) by the engine during this process?
20. One liter of air is cooled at constant pressure until its volume is halved, and then it is allowed to expand isothermally back to its original volume. Draw the process on a *p-V* diagram.
21. In an isothermal process (27 °C), 2 kilomole of an ideal gas is compressed from a volume of 4 litre to 1 litre. Find the work done on the system. (R = 8.3143 J mol^{-1} K^{-1})

22. An automobile engine has an efficiency of 20 % and produces an average of 23 000 J of mechanical work per second during operation. (i) How much heat input is required per second and (ii) how much heat is discharged as waste heat from this engine per second?

CONCEPT MAP

HEAT TRANSFER AND THERMODYNAMICS

Heat Transfer

- heat conduction

$$H_{cond} = k\,A\frac{T_2 - T_1}{l}$$

- heat convection

$$H_{conv} = q\,A\,(T_2 - T_1)$$

- heat radiation

$$H_{rad} = e\,\sigma\,A\,T^4$$

Thermodynamics

Laws of thermodynamics

- zeroth law of thermodynamics
- first law of thermodynamics

$$\Delta U = Q - W$$

- thermodynamic processes

Heat engines

- diesel engine
- gasoline engines
- Carnot heat engine

$$\eta = \frac{W}{Q_H}$$

$$\eta = 1 - \frac{Q_C}{Q_H}$$

$$\eta_C = 1 - \frac{T_C}{T_H}$$

CHAPTER 6

EFFECTS AND CHARACTERISTICS OF SOUND

In this chapter, the range of audibility, infrasound and ultrasound together with their applications will be studied. Echo, reverberation and Doppler effect are also discussed.

Learning Outcomes

It is expected that students will

- examine the propagation of sound.
- identify audible frequencies.
- investigate the infrasound, ultrasound and their applications.
- describe the occurrence of echo.
- examine and explain the Doppler effect.
- differentiate between pitch and loudness of sound.

6.1 PROPAGATION OF SOUND IN A MEDIUM

Sound is a sequence of pressure waves which propagates through compressible media such as air or water. During their propagation, waves can be reflected, refracted or attenuated by the medium. The followings are the properties of media which affect the sound propagation.

- A relationship between density and pressure. This relationship, affected by temperature, determines the speed of sound within the medium.
- The viscosity of the medium. This determines the rate at which sound is attenuated. For many media, such as air or water, attenuation due to viscosity is negligible.

Speed of Sound

Speed of a sound wave can also be affected by the density, temperature, and elasticity of the medium through which the sound waves travel. Hence, the speed of sound will be different for different media.

In a solid medium, the speed of sound is given by,

$$v = \sqrt{\frac{Y}{\rho}} \qquad (6.1)$$

where Y is Young's modulus and ρ is density of the solid material.

In liquids, the velocity of sound is given by a similar expression,

$$v = \sqrt{\frac{B}{\rho}} \qquad (6.2)$$

where B is the bulk modulus and ρ is density of liquid.

In gases, the velocity of sound is given by,

$$v = \sqrt{\frac{\gamma p}{\rho}} \qquad (6.3)$$

where γ is the adiabatic constant of gas, p is pressure in gas and ρ is its density.

Longitudinal sound waves travel faster in solids than they do in liquids than they do in gases.

Young's modulus and the bulk modulus are elasticity properties of the medium. The SI units of Young's modulus and the bulk modulus are N m^{-2}.

Example 6.1 What is the speed of sound in (i) a steel rod (ii) water?

Young's modulus of steel is 20 $\times 10^{10}$ N m^{-2} and density of steel is 7.8 $\times$ 10^3 kg m^{-3}. The bulk modulus of water is 0.22 $\times 10^{10}$ N m^{-2} and density of water is 10^3 kg m^{-3}.

(i) Young's modulus of steel $Y = 20 \times 10^{10}$ N m^{-2}, density of steel $\rho = 7.8 \times 10^3$ kg m^{-3}

The speed of sound in a steel rod $v = \sqrt{\frac{Y}{\rho}}$

$$v = \sqrt{\frac{20 \times 10^{10}}{7.8 \times 10^3}}$$

$$= 5.06 \times 10^3 \text{ m s}^{-1}$$

(ii) The bulk modulus of water $B = 0.22 \times 10^{10}$ N m^{-2}, density of water $\rho = 10^3$ kg m^{-3}.

The speed of sound in water $v = \sqrt{\frac{B}{\rho}}$

$$v = \sqrt{\frac{0.22 \times 10^{10}}{10^3}}$$

$$= 1.48 \times 10^3 \text{ m s}^{-1}$$

Example 6.2 Calculate the velocity of sound in air at STP. The density of air at STP is 1.29 kg m^{-3}. Assume air to be diatomic with $\gamma = 1.4$. Also calculate the velocity of sound in air at 27 °C.

At STP, atmospheric pressure $p = 1.01 \times 10^5$ Pa , temperature $T = 273$ K

Density of air $\rho = 1.29$ kg m^{-3} , $\gamma = 1.4$

The velocity of sound in air is, $v = \sqrt{\frac{\gamma p}{\rho}}$

$$= \sqrt{\frac{1.4 \times 1.01 \times 10^5}{1.29}} = 331.07 \text{ m s}^{-1}$$

At 27 °C , $T_2 = 27 + 273 = 300$ K, $T_1 = 273$ K

Since $v \propto \sqrt{T}$,

$$\frac{v_2}{v_1} = \sqrt{\frac{T_2}{T_1}}$$

The velocity of sound in air at 27 °C is,

$$\frac{v_{300K}}{v_{273K}} = \sqrt{\frac{300}{273}}$$

$$v_{300K} = 331.07\sqrt{\frac{300}{273}}$$

$$= 347.06 \text{ m s}^{-1}$$

Intensity Level of Sound

The level of intensity of a sound is measured relative to a reference intensity value I_0, taken as 10^{-12} W m^{-2} which correspond to the minimum intensity of sound that can be heard (threshold of hearing or minimum audible intensity).

The decibel (dB) is the unit used for measuring sound intensity (intensity level of sound). The intensity level of sound β in dB is defined by Eq. (6.1).

$$\beta\,(\text{dB}) = 10 \log_{10} \frac{I}{I_0} \qquad (6.4)$$

where I is the intensity of sound and I_0 is the reference intensity. For every 10 dB added to the sound level, the loudness of the sound is increased by 10 times. That is 120 dB sound will be 100 times louder than 100 dB sound. Few examples of sound levels (representative average values) are listed in Table 6.1.

Table 6.1 Few examples of sound levels

Sources of sound	Intensity level of sound	
	(dB)	(W m^{-2})
jet plane taking off	120	1
loud music	100	10^{-2}
busy street traffic	70	10^{-5}
people talking	50	10^{-7}
whisper	20	10^{-10}
leaves falling	10	10^{-11}
threshold of hearing	0	10^{-12}

Example 6.3 At a busy street corner, the intensity level of sound is 70 dB. What is intensity of sound in W m^{-2}?

Intensity level of sound $\beta = 70$ dB, threshold of hearing $I_0 = 10^{-12}$ W m^{-2}

$$\beta = 10 \log_{10} \frac{I}{I_0}$$

$$70 = 10 \log_{10} \frac{I}{I_0}$$

$$7 = \log_{10} \frac{I}{I_0}$$

$$\frac{I}{I_0} = 10^7$$

$$I = 10^7 \times 10^{-12} = 10^{-5} \text{ W m}^{-2}$$

Example 6.4 What value of sound intensity in dB increases by a factor of 1000?

$$\frac{I}{I_0} = 1000$$

Intensity level of sound

$$\beta = 10 \log_{10} \frac{I}{I_0}$$

$$= 10 \log_{10} 10^3 = 30 \text{ dB}$$

If the sound intensity is increased by a factor of 1000, the value of sound intensity increases 30 dB.

Sound intensity is a direct measure of the power reaching an eardrum measured in W m^{-2}.

Reviewed Exercise

1. Why is the speed of sound higher in solids than liquids and gases?
2. If the intensity of sound is doubled, by how many decibels does the sound level increase?

Key Words: air pressure, viscosity, Young's modulus, bulk modulus, adiabatic constant, elasticity,

6.2 ECHO

Echoes are produced where a sound wave is reflected from a large and hard surface. An echo is repetition of sound due to the reflection of sound. The echo or the reflected sound can be heard separately from the original sound if the source of the sound is much closer to the observer than to the reflecting surface. Reflection of sound obeys the laws of reflection.

To reduce the effect of echo in buildings, walls are covered up with padding and the floors are covered with rugs and carpets. These materials absorb the incident sound and reduce the amount of sound reflecting off wall and floor surfaces.

Echo can also be used to determine the distance d between a source of sound and the reflecting surface by using formula,

$$v = \frac{2d}{t} \qquad (6.5)$$

where v is the speed of sound and t is the echo time.

Some animals such as bats are known to use **echolocation to navigate and locate their prey as** shown in Figure 6.1. Echolocation is the special ability of **emitting sounds and interpreting the** echoes.

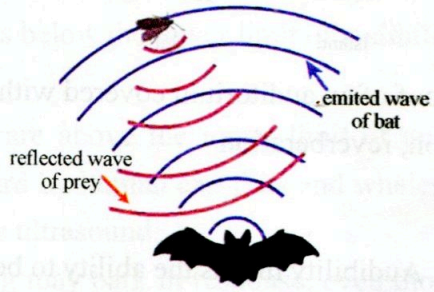


Figure 6.1 Echolocation of bat

Example 6.5 A bat flying at a speed of 20 m s^{-1} towards a stationary moth emits sound wave of 340 m s^{-1}. The bat hears the echoes 0.4 s later. (i) Calculate the distance travelled by the bat in 0.4 s. (ii) Calculate the distance travelled by the sound wave in 0.4 s. (iii) Determine the initial distance between the bat and the moth.

Speed of bat, $v_b = 20$ m s^{-1}, speed of sound, $v_s = 340$ m s^{-1}, $t = 0.4$ s

(i) speed $v = \dfrac{d}{t}$

distance travelled by the bat $d_b = v_b t$

$= 20 \times 0.4 = 8$ m

(ii) distance travelled by sound wave $d_s = v_s t$

$= 340 \times 0.4 = 136$ m

(iii) The sound wave will travel from the initial location of the flying bat to the stationary moth and bounce off as echoes. The echoes travel back to the bat at its position after 0.4 s. Hence, the total distance travelled by the bat and the sound wave is twice that of the initial distance between the bat and the moth.

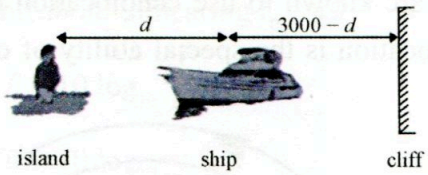
Total distance $= 8 + 136 = 144$ m

Initial distance between the bat and the moth $= \dfrac{\text{total distance}}{2} = \dfrac{144}{2} = 72$ m

A reverberation is a reflection of light or sound waves, or a wide reaching impact of an action. An example of a reverberation is the sound bouncing around in a large speaker.

Reviewed Exercise

1. The sole survivor of a shipwreck swim to an island, which is 3000 m from a vertical cliff. He sees a ship anchored between the island and the cliff. A blast from the ship horn is heard twice with the time elapse of 4 s. Calculate distance between the ship and the survivor. (Speed of sound in air = 330 ms^{-1})



2. Why are the walls and roof of an auditorium covered with sound absorbent materials?

Key Words: echo, echolocation, reverberation

6.3 AUDIBILITY

Audible means able to be heard. Audibility means the ability to be heard or capability of detecting sounds in certain range of frequencies.

The human's ear is capable of detecting sounds only in a certain range of frequencies. The range of frequencies a person can hear is known as the range of audibility. For humans with normal hearing, the lower limit is 20 Hz and the upper limit is about 20 000 Hz. These limits are known as the limit of audibility. The audible range of different mammals are as shown in Figure 6.2.

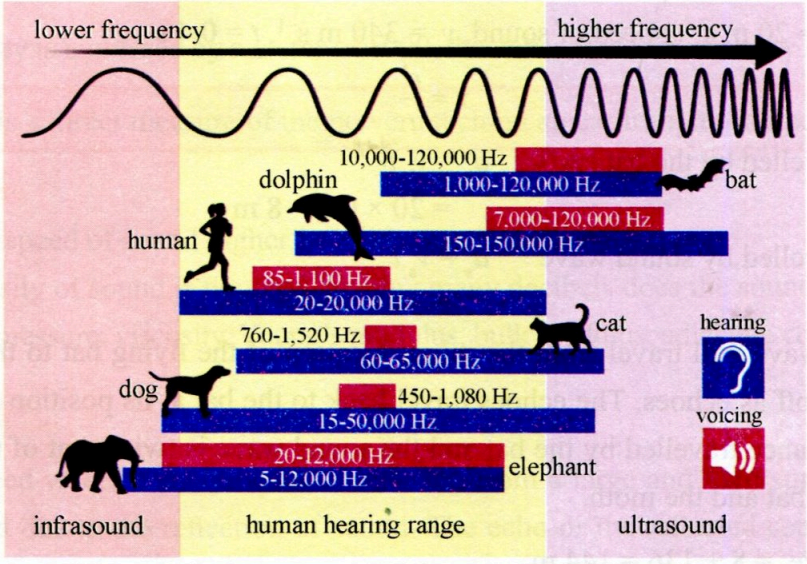


Figure 6.2 Audible range of different mammals [CREDIT: Source from the Internet]

Determination of the Audible Frequency Range by using Frequency Generator

By using frequency generator, the audible frequency range can be determined. Some portable sound frequency generator can create sound with frequencies 0 Hz to about 22 kHz. Tune the sound frequencies gradually from 0 Hz and record the frequency when start hearing the sound, i.e. the lower limit of audibility. Then increase the sound frequency up to about 20 kHz, and record the highest frequency above which the sound cannot be heard, i.e. the upper limit of audibility.

Infrasound

Low frequency sounds which are under the lower limit of audibility (20 Hz) are called infrasound and cannot be heard by human. A vibrating ruler can be seen but not heard its vibrating sound. This is because the sound produced is below the lower limit of audibility of human ear.

Ultrasound

High frequency sounds which are above the upper limit of audibility (20 000 Hz) are called ultrasound which cannot be heard by human ear. Bats and whales use ultrasound in echolocation and most sonar technologies use ultrasound.

If you blow a dog whistle, a dog may bark in response, even though you do not hear any sound. This is because the frequency of the sound produced by the whistle is above the upper limit of audibility of humans but within that of dogs.

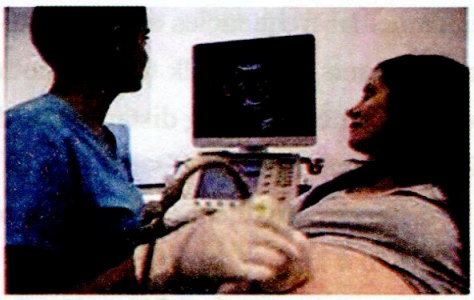
Various Applications of Ultrasound

1. Quality Control

Manufacturers of concrete use ultrasound to check for cracks or cavities. It can also be used to inspect metal pipes and measure the thickness of the wooden boards.

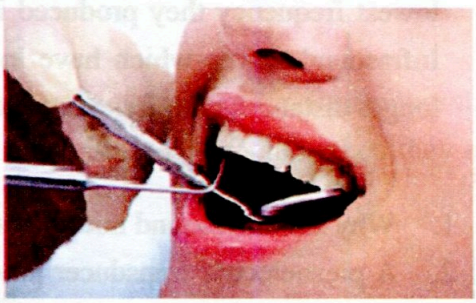
2. Prenatal Ultrasound Test

A prenatal ultrasound test uses high-frequency sound waves, inaudible to the human ear, that are transmitted through the abdomen via a device called a transducer to look at the inside of the abdomen. With prenatal ultrasound, the echoes are recorded and transformed into video or photographic images of the baby.



3. Ultrasonic Scalers in Dentistry

Ultrasonic scalers use ultrasound to remove calculus (plaque) deposits from the teeth. They work mechanically using high frequency ultrasound that effectively blasts calculus. The plaque bacteria and calculus can be flushed from the teeth by a small jet of water or antibacterial mouthwash which is ejected from the top of the scaler.

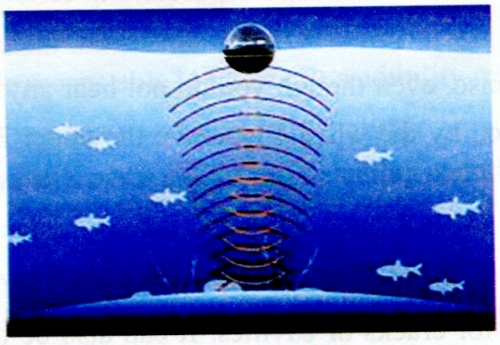


4. Ultrasonic Cleaning

The transmission of high energy ultrasound may result in the certain of cavitation bubbles. Cavitation bubbles are created at sites of rarefaction. These cavitation bubbles may displace contaminant from surfaces. This effect also allows fresh chemicals to come into contact with the contaminant remaining on the surface to be removed. Ultrasonic cleaning is especially effective in the cleaning of irregular surfaces or internal cavities and passageways (such as medical ultrasonic cleaner, industrial ultrasonic cleaner).

5. Sonar

The sonar (sound navigation and ranging) is a type of technology that works based on echoes. It is used by the ships for navigation at sea and to detect the position of other vessels. For example, the depth of the sea or the position of the shoals of fish can be found by sending out a signal (a pulse of sound) from a transducer (transmitter) and noting the time interval for the reflected signal (the echo) to be received as shown in Figure 6.3.



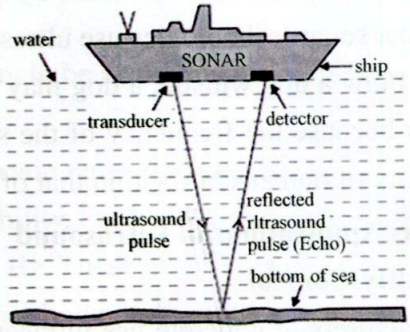


Figure 6.3 Illustration of sonar

Example 6.6 Elephants communicate over a long distance (up to 4 km) using sound with the fundamental frequencies as low as 15 Hz. A rhinoceros uses frequencies down to 10 Hz. (i) Why are elephants able to talk to one another over such a long distance? (ii) Can rhinoceros be able to communicate over long distances further than 4 km? Why?

(i) Infrasound waves ($f < 20$ Hz) have long wavelengths, and are thus able to travel over long distances. So, elephants are able to talk using frequencies as low as 15 Hz over long distances (up to 4 km).

(ii) Rhinoceros can be able to communicate over long distances further than 4 km because the lowest frequency they produced is 10 Hz which is less than 15 Hz and longer wavelength. Infrasound waves which have lower frequency and longer wavelengths are able to travel longer distances with less attenuation.

Reviewed Exercise

1. Why can infrasound not be heard by humans?
2. A piezoelectric transducer produces ultrasound at a frequency of 10 MHz. Calculate the wavelength of this ultrasound in blood if the speed of ultrasound in blood is 1560 ms^{-1}.

Key Words: range of audibility, infrasound, ultrasound

6.4 THE DOPPLER EFFECT

The Doppler effect or Doppler shift is the change in frequency of a wave related to an observer due to the relative motion between the source and observer. It is named after the Austrian physicist Christian Doppler, who described the phenomenon in 1842. Usually, only the motion of the source

and the observer directly towards or away from each other in the reference frame in which the wave medium is considered at rest. A common example of Doppler shift is the change of frequency (pitch) heard when a vehicle sounding a horn approach and recedes from an observer as shown in Figure 6.4.

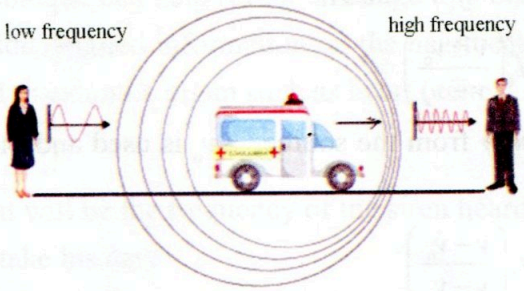


Figure 6.4 Illustration of Doppler effect

Doppler Equation

There are three cases we encounter in Doppler effect:

(i) The observer is stationary and the source is moving towards or away from the observer.

(ii) The source is stationary and the observer is moving towards or away from the source.

(iii) Both the source and observer are moving relative to each other.

Without going into any detail, the Doppler equation can be expressed in general form as

$$f_o = f_s\left(\frac{v \pm v_o}{v \mp v_s}\right) \qquad (6.6)$$

where f_o = frequency heard by the observer

f_s = frequency produced by the source

v = speed of sound in the medium

v_o = speed of the observer

v_s = speed of the source

Note that the appropriate signs must be taken into account in using this equation. Remember that when the source and observer are approaching relative to each other $f_o > f_s$; and when they are receding relative to each other $f_o < f_s$. Hence,

$+ v_o$ is used when the observer moves towards the source,

$- v_o$ is used when the observer moves away from the source,

$- v_s$ is used when the source moves towards the observer, and

$+ v_s$ is used when source the moves away from the observer.

When observer and source are stationary, $v_o = v_s = 0$.

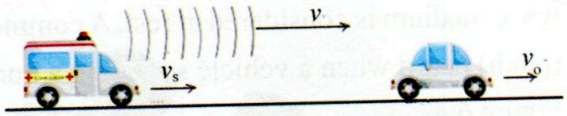
Example 6.7 An ambulance travelling at 44 m s^{-1} approaches a car heading in a same direction at the speed of 28 m s^{-1}. The ambulance driver has a siren sounding at 550 Hz. At what frequency does the driver of the car hear the siren? (Speed of siren sound = v = 330 m s^{-1})

Speed of the car (observer) $v_o = 28 \text{ m s}^{-1}$

Speed of the ambulance (source) $v_s = 44 \text{ m s}^{-1}$

Speed of the siren sound $v = 330 \text{ m s}^{-1}$

Frequency of the siren sound $f_s = 550 \text{ Hz}$



We have $$f_o = f_s\left(\frac{v \pm v_o}{v \mp v_s}\right)$$

Since the observer moves away from the source, $-v_o$ is used and the source moves towards the observer, $-v_s$ is used.

Therefore, $$f_o = f_s\left(\frac{v - v_o}{v - v_s}\right)$$

$$= 550\left(\frac{330-28}{330-44}\right) = 580.8 \text{ Hz}$$

Application of Doppler Effect

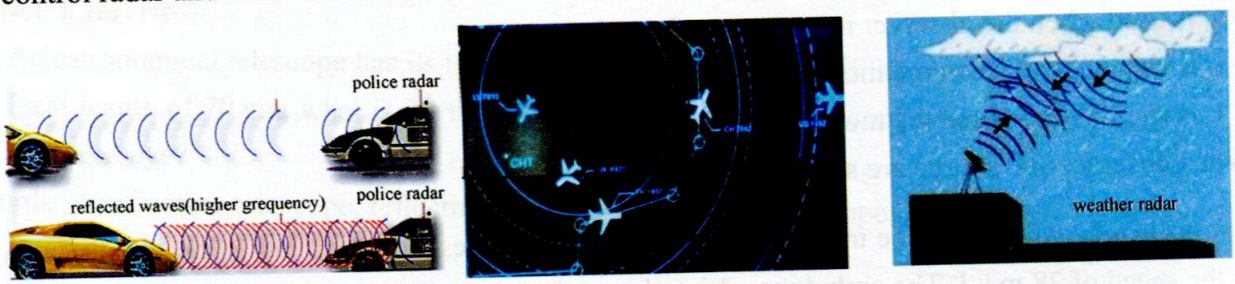
Doppler effect is observed not only in sound but also in electromagnetic radiations (radio wave, microwave, light, etc.). Some applications of Doppler effect in a variety of fields is described below.

1. Red Shift

The wavelength of light from a star observed from the earth shifts due to Doppler effect towards the red end of the spectrum (lower frequency, longer wavelength) when the earth and the star are moving away from each other. This phenomenon is called red shift and the velocity of the astronomical body moving away from the earth can be determined from the fractional change in wavelength ($\frac{\Delta\lambda}{\lambda} = \frac{v}{c}$). If the star were approaching the earth the spectral lines would shift towards blue end.

2. Radar

Radar (radio detection and ranging) is a form of echolocation that uses electromagnetic waves instead of sound but otherwise the concept is similar. The Doppler effect is used in some types of radar, to measure the velocity of detected objects. Some examples are police radar, air traffic control radar and weather radar as shown in Figure 6.5 (a), (b) and (c).



(a) (b) (c)

Figure 6.5 (a) Police radar, (b) air traffic control radar and (c) weather radar

3. Medical Applications

Doppler ultrasound is a technique that is used to examine blood flow. Blood flow rates can be found by measuring Doppler shift in frequency of ultrasound reflected by red blood cell (known as angiodynography). This technique can help reveal blockage to blood flow, show the formation of plaque in arteries and provide detailed information on the heartbeat of the fetus during labor and delivery. It can also be used to examine organ such as heart (echocardiography) and kidney.

Reviewed Exercise

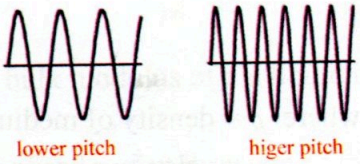
1. In example 6.7, what will be the frequency of the siren heard by the driver of the car when the ambulance overtake his car?
2. A monorail approaches a platform at a speed of 50 m s^{-1} which it blows its whistle. A man standing on the platform hears the whistle with frequency 261 Hz. There is no wind and temperature is chilly 0 °C . What is the observed frequency of the whistle when the train is at rest? (The speed of sound in air at 0 °C is 332 m s^{-1})

Key Words: Doppler effect, red shift, radar, Doppler ultrasound

6.5 PROPERTIES OF SOUND WAVES

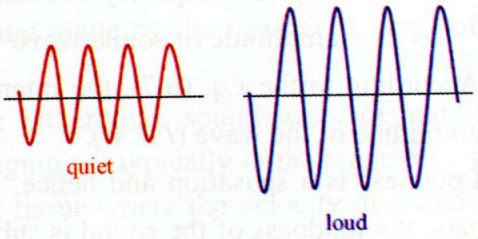
Pitch

The pitch of the note (how high it is) is determined by its frequency. The greater the frequency, the higher is the pitch.



Loudness

The loudness of the note depends upon the amplitude of the wave that produces it. The greater the amplitude, the louder the note, because more energy is used to produce a larger amplitude. The amplitude of the wave on the right is twice that of the wave on the left. So, the loudness is four times greater.



Quality or Timbre

Quality of a sound is the characteristic which distinguishes the two sounds of the same loudness and same pitch but emitted by two different instruments.

In Figure 6.6, the notes have the same frequency and hence the same pitch. However, they sound differently because of their different waveforms. That is, they differ in quality or timbre. The notes produced by the violin richer than others because it has more overtones. Notes of the same pitch

played upon different musical instruments are distinguished from each other by their quality. That is, the timbre of a sound depends on its waveform, which varies with the number of overtones, or harmonics.

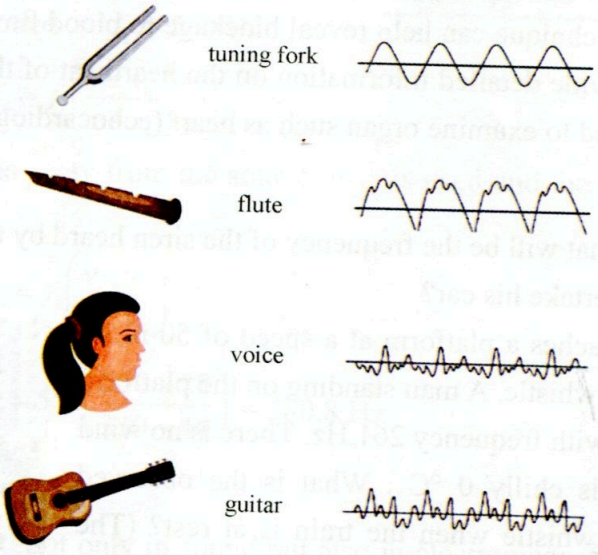


Figure 6.6 Different quality of sound produced by different musical instruments same pitch

Intensity and loudness of sound related to amplitude

The intensity of the sound at a place is the energy transmitted per second per unit area held normally to direction of sound and can be expressed as

$$I = \frac{1}{2}\rho v\omega^2 A^2 \qquad (6.7)$$

where ρ = density of medium

v = velocity of sound wave

ω = angular frequency of sound wave

A = amplitude of sound wave

According to the Eq. (6.7), the intensity of a wave is directly proportional to the square of the amplitude of the wave ($I \propto A^2$).

Loudness is a sensation and hence, unlike intensity, it depends on the observer. To the human ears, the loudness of the sound is subjective. For a particular volume of sound, some may find it loud, while others finding it soft. Normally the greater the intensity, the greater the loudness of the sound is.

Example 6.8 Calculate the displacement of air molecules (amplitude) for a sound having a frequency of 1000 Hz at the threshold of hearing. The speed of sound in air is 343 m s^{-1} at 20 °C and density of air is 1.29 kg m^{-3}.

Frequency of sound f = 1000 Hz, density of air ρ = 1.29 kg m^{-3}

Angular frequency $\omega = 2\pi f = 2 \times 3.142 \times 1000 = 6284$ rad s^{-1}

Speed of sound in air $v = 343\ m\ s^{-1}$

At the threshold of hearing, intensity of sound $I = 1 \times 10^{-12}\ W\ m^{-2}$

$$I = \frac{1}{2}\rho v\omega^2 A^2$$

$$10^{-12} = \frac{1}{2} \times 1.29 \times 343 \times (6284)^2 A^2$$

$$A = .1.07 \times 10^{-11}\ m$$

Note that the amplitude of air molecules is very small.

Reviewed Exercise

- In terms of loudness and pitch, how would you compare the sounds made by a mosquito flying near your ear and the cracking of a toad?

Key Words: pitch, loudness, quality of sound (timbre)

SUMMARY

Audibility means the ability to be heard or capability of detecting sounds in certain range of frequencies.

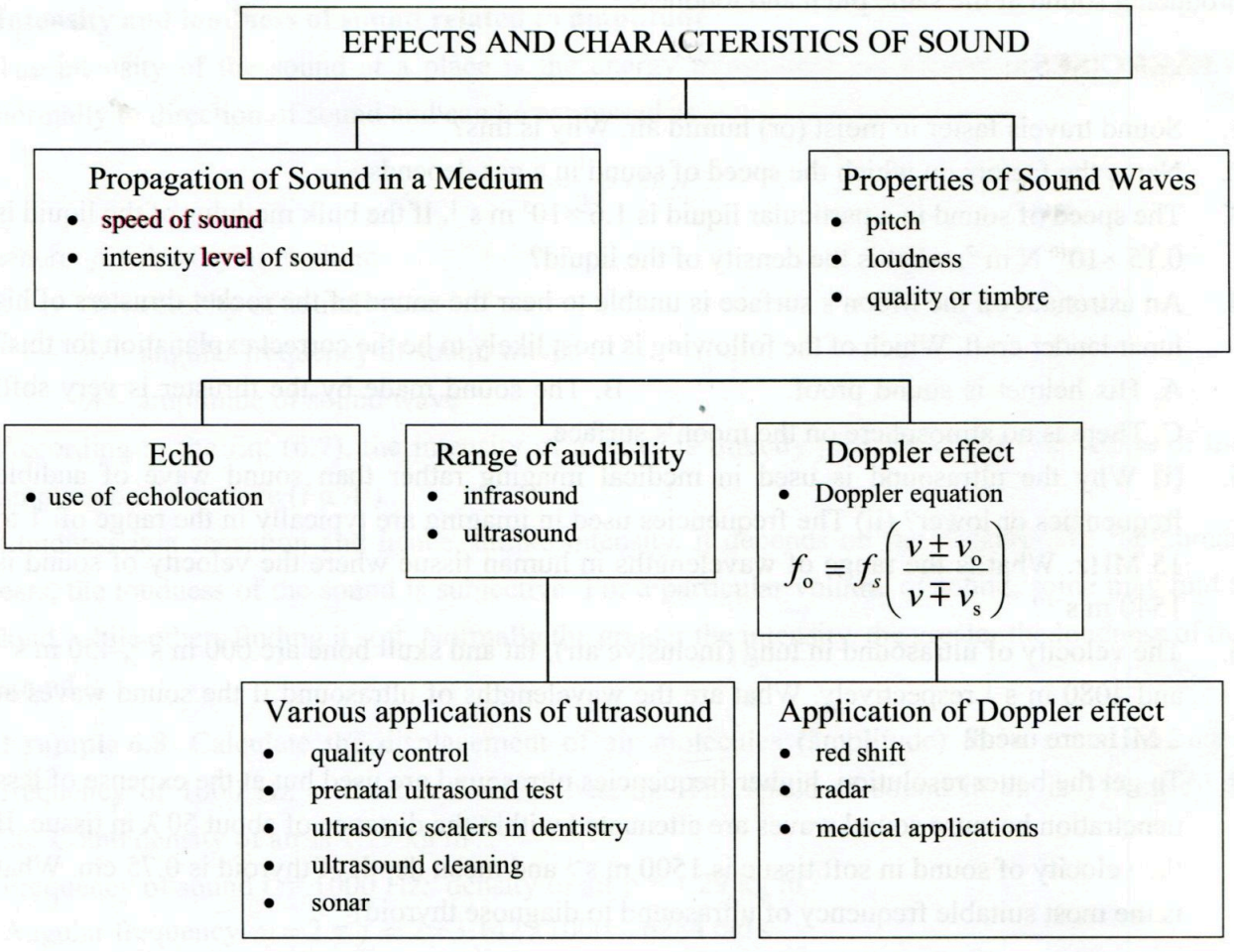
Timbre is the quality of sound which allows us to distinguish between different sound sources producing sound at the same pitch and loudness.

EXERCISES

1. Sound travels faster in moist (or) humid air. Why is this?
2. Name the factors on which the speed of sound in a gas depends.
3. The speed of sound in a particular liquid is $1.6 \times 10^3\ m\ s^{-1}$. If the bulk modulus of the liquid is $0.15 \times 10^{10}\ N\ m^{-2}$, what is the density of the liquid?
4. An astronaut on the Moon's surface is unable to hear the sound of the rocket thrusters of his lunar lander craft. Which of the following is most likely to be the correct explanation for this?
   A. His helmet is sound proof. B. The sound made by the thruster is very soft.
   C. There is no atmosphere on the moon's surface.
5. (i) Why the ultrasound is used in medical imaging rather than sound wave of audible frequencies or lower? (ii) The frequencies used in imaging are typically in the range of 1 to 15 MHz. What is the range of wavelengths in human tissue where the velocity of sound is $1540\ m\ s^{-1}$?
6. The velocity of ultrasound in lung (inclusive air), fat and skull bone are $600\ m\ s^{-1}$, $450\ m\ s^{-1}$ and $4080\ m\ s^{-1}$ respectively. What are the wavelengths of ultrasound if the sound waves at 2 MHz are used?
7. To get the better resolution, higher frequencies ultrasound are used but at the expense of less penetration because sound waves are attenuated within the distance of about 50 λ in tissue. If the velocity of sound in soft tissue is $1500\ m\ s^{-1}$ and mean depth of thyroid is 0.75 cm. What is the most suitable frequency of ultrasound to diagnose thyroid?

8. Define echo. Give its two applications.
9. Why is the intensity of an echo less than that of the original sound?
10. Elephants and whales are known to use infrasound to communicate over very large distances. What are the advantages of infrasound for long distance communication?
11. A sonar echo returns to a submarine 1.2 s after being emitted. What is the distance to the object creating the echo? (Assume that the submarine is in the ocean, not in fresh water and the velocity of sound in water is 1500 m s^{-1}.)
12. (i) What frequency is received by a person watching an oncoming ambulance moving at 110 km h^{-1} and emitting a steady 800 Hz sound from its siren? The speed of sound on this day is 345 m s^{-1}. (ii) What frequency does she receive after the ambulance has passed?
13. A 5000 Hz sound wave is emitted by a stationary source. This sound reflects from an object moving 3.5 m s^{-1} toward the source. What is the frequency of the wave reflected by the moving object as detected by a detector at rest near the source?
14. Is there a Doppler shift if the source and observer move in the same direction, with the same velocity? Explain.

CONCEPT MAP



CHAPTER 7

APPLICATIONS OF LIGHT AND OPTICAL INSTRUMENTS

Many different types of lenses play an important part in our daily life. Spectacles, cameras, microscopes and telescopes are useful optical instruments employing lenses. The function and formation of images in these instruments are easily understandable with the help of ray diagrams. In this chapter, we will study optical instruments and their applications.

Learning Outcomes

It is expected that students will

- investigate the features of eye and camera.
- explain the function of microscopes and telescopes.
- describe the image created by these instruments and calculate their magnifications.
- study the application of light in instruments such as laser, fibre optics and spectrometer.
- understand the photometry optics.

7.1 THE CAMERA

A camera is an optical instrument that forms and records an image of an object. The image may be recorded on film (or) it may be detected by an electronic sensor that stores the image digitally. Regardless of how the image is recorded, all cameras form images in the same basic way. The essential features of a simple camera are shown in Figure 7.1.

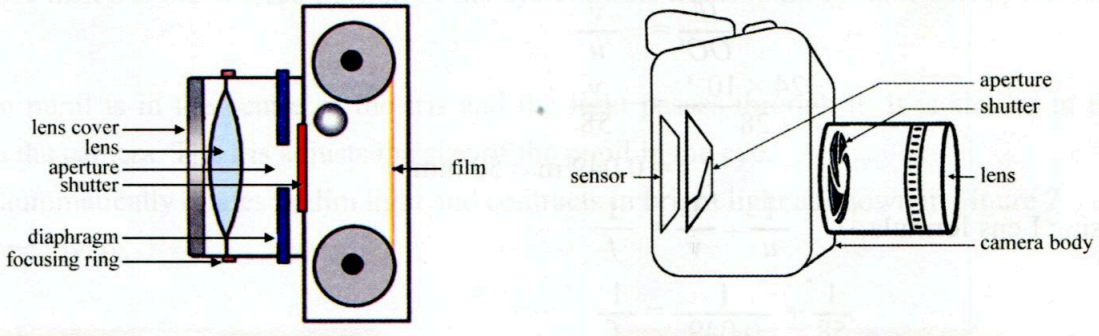


Figure 7.1(a) Camera with film (b) camera with sensor

The essential features of camera are expressed as follows.

Convex lens: A convex lens focuses the light from the object to form a real inverted image on the film or electronic sensor.

Diaphragm: A diaphragm is placed behind the convex lens to control the amount of light energy falling on the film.

Aperture: The aperture is a circular opening through which the light passes. It can be made wide or narrow by adjusting the diaphragm.

The size or the diameter of the aperture is expressed in terms of *f*-numbers. The *f*-number of a camera is the ratio of the focal length of the lens to the diameter of aperture. The smaller the *f*-number, the wider the aperture is. When taking a photograph in bright light a large *f*-number (narrow aperture) must be used and when taking a photograph in dim light a small *f*-number (wide aperture) must be used.

Shutter: A shutter controls the time during which light enters the camera through the lens. The shutter speed is expressed in terms of the exposure time. The shorter the exposure time, the higher the shutter speed is.

Focusing Ring: A camera consists of a focusing ring by which the distance between the lens and the film can be adjusted to obtain a sharp image on the film.

Film: The film at the back of the camera is actually a screen which receives the image. The film consists of a thin layer of chemical compounds of silver. Chemical reaction occurs when light energy falls on the film and as a result the image of an object is formed on the film. In a digital camera, the film is replaced by an electronic sensor. It is made up of millions of tiny pixels which response to light.

Example 7.1 A nature photographer wishes to photograph a 28 m tall tree from a distance of 58 m. What focal length of lens should be used if the image is to fill the 24 mm height of the film?

Object height $OO' = 28$ m, object distance $u = 58$ m

Since the image formed by camera lens is real and inverted, image height $II' = -24$ mm

$$\frac{II'}{OO'} = -\frac{v}{u}$$

$$\frac{-24 \times 10^{-3}}{28} = -\frac{v}{58}$$

$$v = 0.049 \text{ m} \approx 50 \text{ mm}$$

Using Lens formula, $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

$$\frac{1}{58} + \frac{1}{0.049} = \frac{1}{f}$$

$$f = 0.049 \text{ m} \approx 50 \text{ mm}$$

It is noted that for a distant object its image is formed at focus of the lens.

Reviewed Exercise

- Why must a camera lens be moved further from the film to focus on a closer object?

Key Words: convex lens, film, diaphragm, aperture, shutter, exposure time

7.2 HUMAN EYE

Human eye consists of sense organs which are capable of receiving visual images and sending the information to the brain. The human eye is very similar to a camera in many aspects. The optical features of the eye are shown in Figure 7.2.

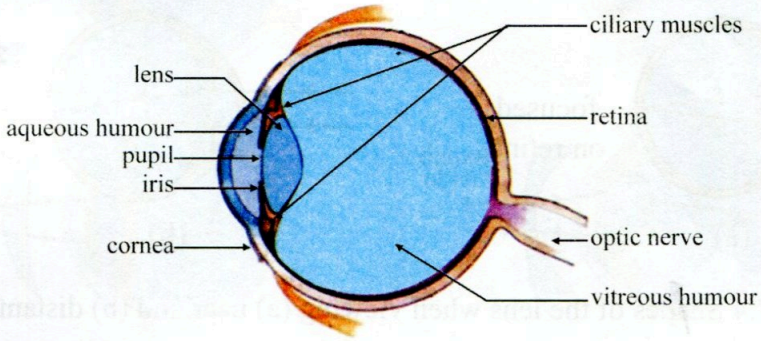


Figure 7.2 Optical features of the eye

The main features of human eye are described as follows.

Eyelid: Since the eyelid opens and closes the eye, it is similar to the lens cover of the camera.

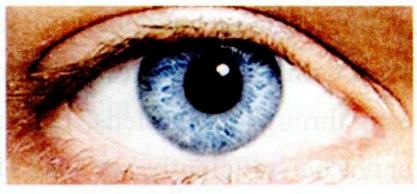
Cornea: The front portion of the eye is covered by a tough, transparent membrane called the cornea.

Aqueous humour and vitreous humour: The region between the cornea and the eyelens is filled with a transparent liquid called the aqueous humour. The region between the eyelens and the retina contains a transparent liquid called the vitreous humour.

Iris: The iris which is the colored portion of the eye controls the amount of light energy entering the eye.

Pupil: The pupil is in the centre of the iris and the light passes through it. It is similar to the aperture in the camera. The iris adjusts the size of the pupil in the eye.
The pupil automatically dilates in dim light and contracts in bright light as shown in Figure 7.3 (a) and (b) respectively.

(a)



(b)

Figure 7.3 Behavior of the pupil in (a) dim light and (b) bright light

Ciliary muscles: The ciliary muscles are attached to the eyelens. The ciliary muscles can vary the thickness of the eyelens, thus alter the focal length of the lens to obtain a sharp image on the retina. When near objects are viewed, the ciliary muscles contract and permit the eyelens to form a more convex shape as ahown in Figure 7.4 (a). When distant objects are viewed, the ciliary muscles relax and the eyelens becomes less convex as shown in Figure 7.4 (b).

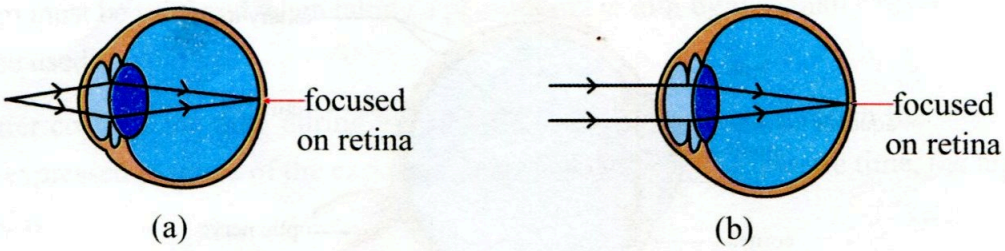


Figure 7.4 Shapes of the lens when viewing (a) near and (b) distant objects

Retina: The retina is the light sensitive area of the cells, at the back of the eye. The retina consists of rods and cones. The yellow spot which is the most light-sensitive spot on the retina is used for viewing very small objects. Since the yellow spot consists of cones exclusively, the small objects can be seen clearly. There is a spot on the retina which has no rods or cones. When an image is formed at this spot it cannot be seen at all. This spot is called the blind spot where the optic nerve enters the eye. The light energy falling on the retina is transmitted to the brain by means of the optic nerve.

The cornea, the aqueous humour, the eyelens and the vitreous humour are refracting substances. The image of an object formed on the retina by these four substances is real, inverted and smaller than the size of the object.

Common Defects of Vision

The common defects of vision are farsightedness (hyperopia) and nearsightedness (myopia). Farsightedness is a defect of vision in which distant objects are seen clearly, but objects nearby are blurry. On the other hand, nearsightedness is a defect of vision in which near objects are seen clearly, but objects far away are blurry. These defects are caused by the ciliary muscles which cannot contract or relax sufficiently or by wrongly shaped eyeballs.

For a person with farsightedness, the rays from near objects intersect at a point behind the retina after passing through the eyelens. This is due to an eyeball which is too short or due to the eyelens which cannot be sufficiently convex. This defect can be corrected by the use of spectacles with converging lens or convex lens as shown in Figure 7.5 (a).

For a person with nearsightedness, the rays from distant objects intersect at a point in front of the retina after passing through the eyelens. This is due to an eyeball which is too long or due to

the eyelens which is more convex than is necessary. This defect can be corrected by the use of spectacles with diverging lens or concave lens as shown in Figure 7.5 (b).

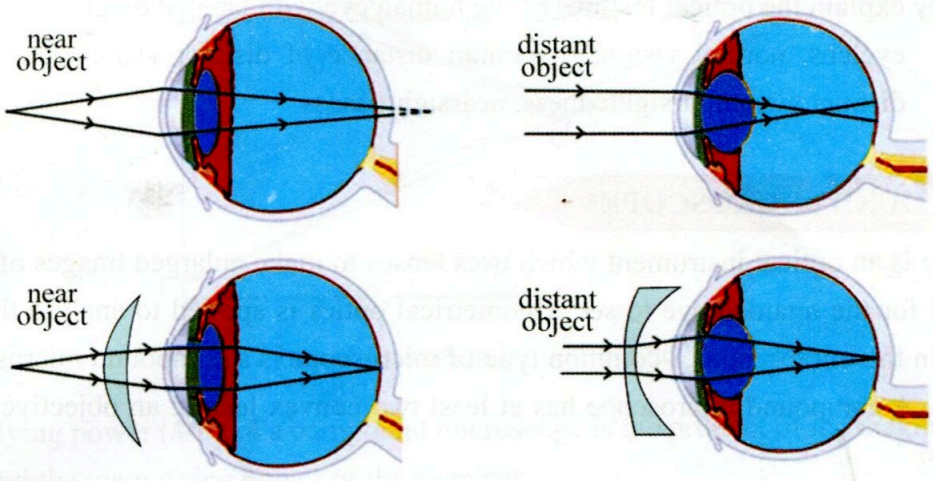


Figure 7. 5 (a) Farsightedness and correction (b) nearsightedness and correction

People with normal vision can see an object in greatest detail when it is placed at a certain distance from the eye. This distance, which is known as the least distance of distinct vision is about 25 cm. People with normal vision can see clearly objects which are at infinity. Thus, for normal vision the maximum distance of distinct vision is infinity. When objects at infinity are viewed, the eyes are relaxed and such objects can be viewed conveniently for a long time.

Example 7.2 Calculate the power of the eyelens for viewing object at the longest and shortest distances possible with normal vision, assuming eyelens to retina distance of 2 cm.

For normal vision, the image must be on the retina, $v = +2$ cm $= +0.02$ m (The image formed on the retina is real.)

(i) Object at the maximum (longest) distance of distinct vision, $u \approx \infty$

Using lens formula, $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$

Since, $P = \frac{1}{f(\text{m})}$, $P = \frac{1}{u} + \frac{1}{v}$

$$= \frac{1}{\infty} + \frac{1}{0.02} = 50 \text{ D}$$

(ii) Object at the least (shortest) distance of distinct vision, $u = 25$ cm $= 0.25$ m

$$P = \frac{1}{u} + \frac{1}{v}$$

$$= \frac{1}{0.25} + \frac{1}{0.02} = 54 \text{ D}$$

Reviewed Exercise

1. What are meant by farsightedness and nearsightedness and how can they be corrected?
2. Briefly explain the optical features of the human eye with labeled diagram.

Key Words: eyelens, normal vision, maximum distance of distinct vision, least distance of distinct vision, farsightedness, nearsightedness

7.3 COMPOUND MICROSCOPE

A microscope is an optical instrument which uses lenses to make enlarged images of objects that are too small for the unaided eye to see. Geometrical optics is applied to analyse the formation of an image in the microscope. A common type of microscope is a compound microscope shown in Figure 7.6. A compound microscope has at least two convex lenses: an objective lens and an eyepiece lens.

The lens nearer the object is called the objective and the lens through which the final image is viewed is called the eyepiece. Figure 7.7 shows the formation of magnified image of a very small object by a compound microscope. O represents the objective of short focal length. When the compound microscope is used, the object should be placed between F_o and $2F_o$, where F_o is the focus of the objective. The objective produces an image which is real, inverted and larger than the size of the object. This image is between the eyepiece E and its focus F_e. The eyepiece magnifies this image and the final image formed is virtual and on the same side as the object. The final image can be seen clearly when the eye is very close to the eyepiece.

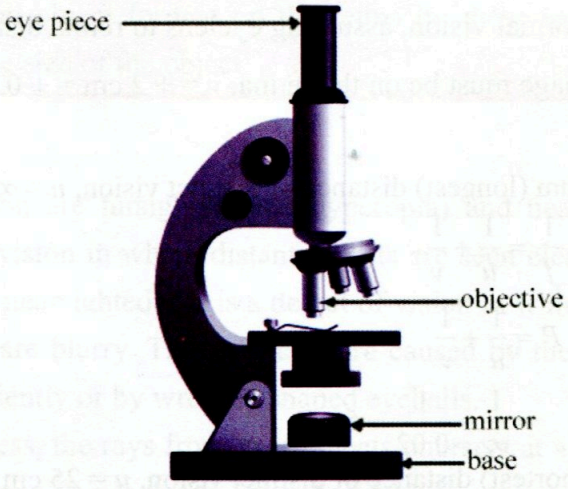


Figure 7.6 Compound microscope

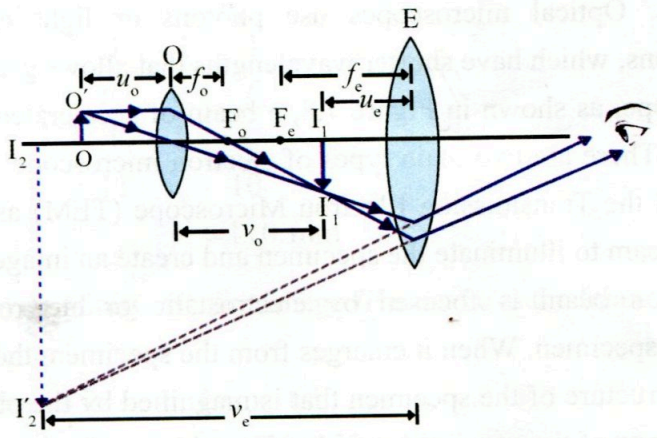


Figure 7.7 Image formation by compound microscope

The magnifying power (*MP*) of a compound microscope is the product of the magnification of the objective and the magnifying power of the eyepiece.

$$MP = \frac{v_o}{u_o} \times \frac{25}{u_e} \qquad (7.1)$$

The separation of the lenses is, $L = v_o + u_e$

(i) When the final image is formed at infinity (for normal setting), $u_e = f_e$.

$$MP = \frac{v_o}{u_o} \times \frac{25}{f_e} \qquad (7.2)$$

(ii) When the final image is formed at a distance of 25 cm from the eyepiece,

$$u_e = \frac{25 f_e}{25 + f_e}$$

Therefore,
$$MP = \frac{v_o}{u_o}\left(\frac{25}{f_e} + 1\right) \qquad (7.3)$$

For a large magnifying power of a compound microscope both the objective and the eyepiece should have short focal lengths. Also, the focal length of the objective should be shorter than that of the eyepiece. Compound microscopes have the magnifying power of several hundreds. As the magnifying power increases, the sharpness of images is reduced. Compound microscopes are used to study living cells and for regular use when magnification is enough.

Electron Microscope

Electron microscope is a microscope with high magnification and resolution, employing electron beams in place of light and using electron lenses. Optical microscopes have a maximum magnifying power of 1000 compared to the better magnification of the electron microscope that can reach 1 000 000. Electron microscopes deliver a more detailed image compared to optical microscopes.

Optical microscopes use simple lenses, whereas electron microscopes use electrostatic or

electromagnetic lenses. Optical microscopes use photons or light energy, while electron microscopes use electrons, which have shorter wavelengths that allows greater magnification.

In an electron microscope, as shown in Figure 7.8, a beam of accelerated electrons is used as a source of illumination. There are two main types of electron microscope: the Scanning Electron Microscope (SEM) and the Transmission Electron Microscope (TEM) as shown in Figure 7.8. They uses an electron beam to illuminate the specimen and create an image.

The accelerated electron beam is focused by electrostatic or electromagnetic lenses, and transmitted through the specimen. When it emerges from the specimen, the electron beam carries information about the structure of the specimen that is magnified by the objective lens system of the microscope. The image of the specimen may be viewed by projecting the magnified electron image onto a fluorescent viewing screen. Alternatively, the image can be recorded by a photographic film or by a digital camera. Electron microscopes are used to investigate the ultrastructure of a wide range of biological and inorganic specimens, metals, and crystals.

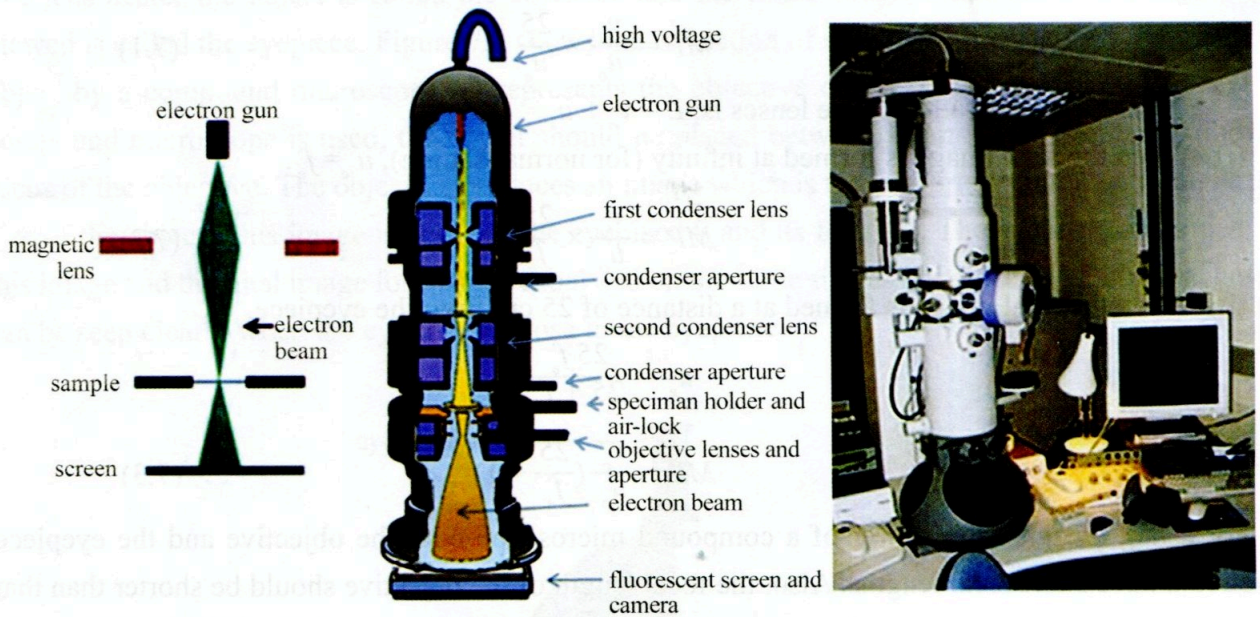


Figure 7.8 Transmission electron microscope (a) basic principle (b) internal structure (c) photograph

Example 7.3 A compound microscope consists of two thin converging lenses. The focal length of the objective is 10 mm and that of the eyepiece is 20 mm. If an object is placed 11 mm from the objective, the instrument produces an image at infinity. Calculate the separation of the lenses and the magnifying power of the instrument.

Object distance from the objective, $u_o = 11$ mm

Focal length of the objective, $f_o = 10$ mm, focal length of the eyepiece, $f_e = 20$ mm

Distance of image v_o produced by the objective is,

$$\frac{1}{u_o}+\frac{1}{v_o}=\frac{1}{f_o}$$

$$\frac{1}{11}+\frac{1}{v_o}=\frac{1}{10}$$

$$v_o = 110 \text{ mm}$$

Since the final image is at infinity, the first image II' must be at the focus of the eyepiece.

Therefore $u_e = f_e = 20$ mm

The separation of the lenses is $L = v_o + f_e$

$= 110 + 20 = 130$ mm

Since the final image is at infinity, the magnifying power is

$$MP = \frac{v_o}{u_o} \times \frac{25}{f_e}$$

$$= \frac{110}{11} \times \frac{250}{20} = 125$$

Reviewed Exercise

1. The final image produced by the compound microscope cannot be projected. Why?
2. A compound microscope 17 cm long has an eyepiece with a focal length of 2.5 cm and objective with focal length of 0.25 cm. What is the magnifying power for normal setting? (Hint: The magnified image of the objective lens forms at the focal point of the eyepiece for normal setting)

Key Words: electron microscope, objective, magnifying power, eyepiece, scanning electron microscope, transmission electron microscope

7.4 TELESCOPE

An optical instrument used for viewing distant objects is called a telescope. Two types of telescope are astronomical telescope and terrestrial telescope.

Astronomical Telescope

The refracting astronomical telescope consists of an objective of long focal length and an eyepiece of short focal length. Both lenses are convex lenses. Galileo invented a refracting telescope.

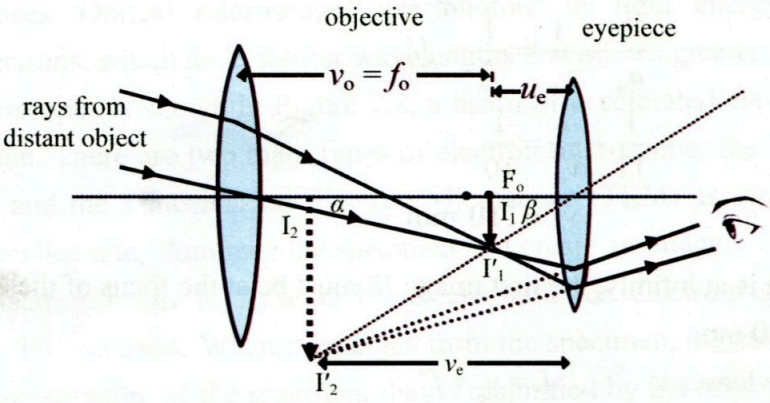


Figure 7.9 Ray diagram of an astronomical telescope

A real and inverted image of a distant object is formed at the focus of the objective. This image is within the focal length of the eyepiece. Hence, the final image formed by the eyepiece is virtual, larger than the first image and inverted with respect to the distant object as shown in Figure 7.9. However, this does not cause any problem in viewing planets, stars and moons.

The magnifying power (*MP*) of the refracting astronomical telescope is defined as the ratio of the angle β subtended by the image seen by the eye through the optical instrument to the angle α subtended by the object as seen without instrument.

$$MP = \frac{\beta}{\alpha} \tag{7.4}$$

From Figure 7.9, $\beta = \tan\beta = \frac{II'}{u_e}$, $\alpha = \tan\alpha = \frac{II'}{v_o}$

Therefore, $$MP = \frac{v_o}{u_e} \tag{7.5}$$

When the final image is formed at infinity, $v_o = f_o$, $u_e = f_e$

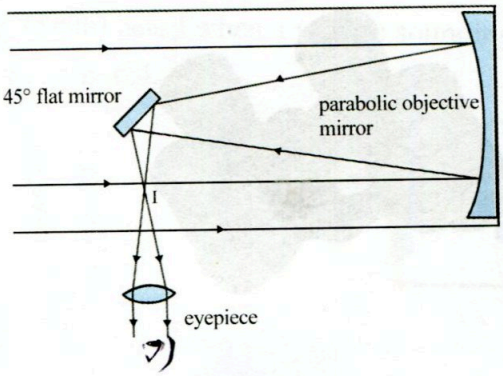
$$MP = \frac{f_o}{f_e} \tag{7.6}$$

The separation of the lenses $L = f_o + f_e$

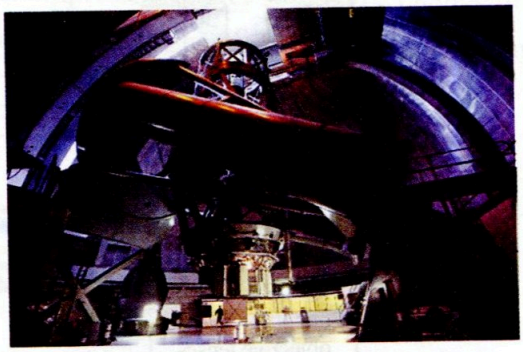
According to E.q (7.6) for a high magnifying power of the refracting astronomical telescope, an objective of long focal length and an eyepiece of short focal length are required. To obtain the sharp images of distant objects the objective should have a large diameter. In practice, it is difficult to make a perfect convex lens of large diameter free from defects so that a concave mirror instead of a convex lens is used as an objective. Most telescopes used at the astronomical observatories in the world are reflecting telescopes with concave mirror used as the objective.

The parallel rays from a distant object are reflected from the concave mirror and they are allowed to fall on a plane mirror. Then, the reflected rays from the plane mirror enter the eye after refraction through the eyepiece as shown in Figure 7.10 (a). The eyepiece is, in fact, a magnifying glass which magnifies the image.

The reflecting telescope was first suggested by Newton and this arrangement is called the Newtonian design (Hale telescope).



(a)

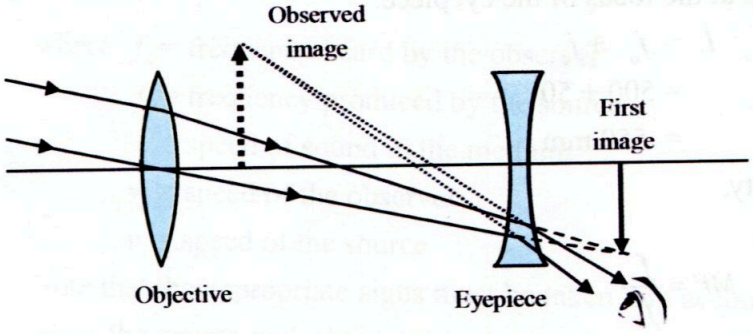


(b)

Figure 7.10 (a) Ray diagram of reflecting astronomical telescope
(b) Hale telescope high resolution stock

Terrestrial Telescope

It is not convenient to use the astronomical telescope for viewing objects on the earth since the final image produced is inverted. The astronomical telescope should be modified to produce a final image which is erect. In the terrestrial telescope a concave lens, instead of a convex lens, is used as an eyepiece to produce a final image which is erect as shown in Figure 7.11(a).



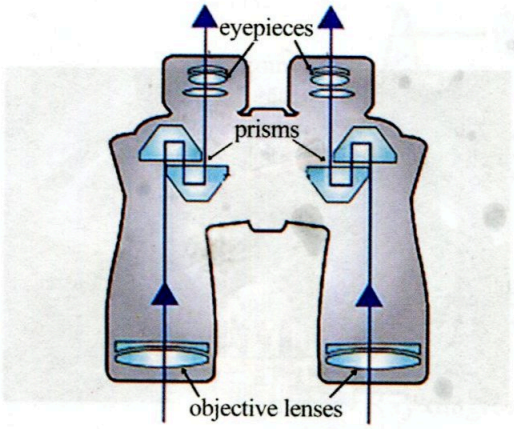
(a)

(b)

Figure 7.11 Terrestrial telescope (a) ray diagram (b) photograph

Since the terrestrial telescope of this type is very long as shown in Figure 7.11(b), it cannot be used conveniently. Total reflecting prisms (90°- 45°- 45° prisms) are used in binoculars to reduce the length of the terrestrial telescope as shown in Figure 7.12. A binoculars is an optical instrument which consists of a pair of telescopes one for each eye, and those telescopes employ the total reflecting prisms. Because of the successive total internal reflections in the prisms, the length of

the telescope is greatly reduced. Binoculars, therefore, are widely used as terrestrial telescopes.



(a)

(b)

Figure 7.12 Binoculars (a) ray diagram (b) photograph

Example 7.4 A telescope consists of two converging lenses: an objective of focal length 500 mm and an eyepiece of focal length 50 mm. When the telescope is in normal adjustment: (i) What is the separation of the lenses? (ii) Where is the final image located? (iii) Is the image erect or inverted? (iv) What is the magnifying power?

Focal length of objective, $f_o = 500$ mm, focal length of eyepiece, $f_e = 50$ mm

When the telescope is in normal adjustment the final image of a distant object is formed at infinity.

Since the object to be viewed is at infinity, its image II' appears at the focus of the objective.

Since the final image is at infinity, II' is at the focus of the eyepiece.

(i) The separation of the lenses $L = f_o + f_e$

$= 500 + 50$

$= 550$ mm

(ii) The final image is located at infinity.

(iii) The image is inverted.

(iv) The magnifying power $MP = \frac{f_o}{f_e}$

$= \frac{500}{50} = 10$

Reviewed Exercise

- An astronomical telescope has its two-lens space 75.2 cm apart. If the objective lens has a focal length of 70 cm, what is the magnifying power of this telescope when viewing with a relaxed eye?

Key Words: refracting telescope, reflecting telescope, terrestrial telescope, binoculars

7.5 LASER

Optics is the branch of physics that studies the nature and properties of light. Light and optics provide various applications in our daily life such as the reflection and refraction of light in mirrors and lenses, and other real-world instruments such as microscope, telescope, lasers, spectrometer and fibre optics.

A laser, acronym for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation, is a source of a narrow beam of intense, monochromatic, coherent light in the visible, infrared or ultraviolet part of the electromagnetic radiation. It produces stimulated emission of photons from excited atoms or molecules. The power in a continuous beam can range from a fraction of a milliwatt to around 20 kilowatts in commercial lasers and up to more than a megawatt in special military lasers.

Lasers are classified into five main types:

- gas lasers
- solid-state lasers
- fiber lasers
- liquid lasers (dye lasers)
- semiconductor lasers (laser diodes)

In practice, a laser contains three essential elements: laser medium, power supply (for pumping process) and optical cavity as shown in Figure 7.13.

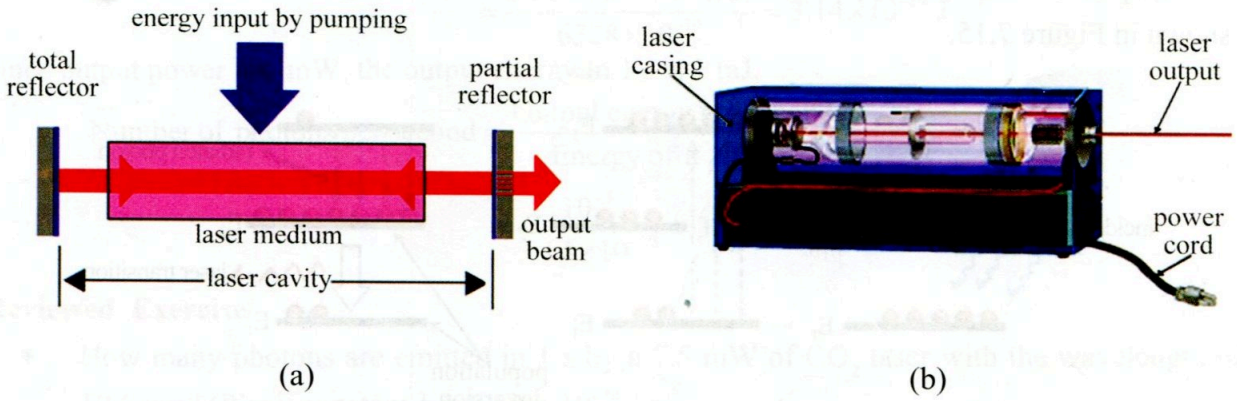


Figure 7.13 A laser system

The laser medium (gain medium) is the material emitting the laser beam. It can be a gas, liquid, crystalline solid or semiconductor crystal. The power supply delivers energy to the laser medium in the form needed to excite it to emit light. The energy is supplied as an electric current, a flash lamp or another laser. This process is called pumping.

The optical cavity (or resonator) concentrates the light to stimulate the emission of laser radiation. Photons are reflected back and forth in the optical cavity increasing the probability of stimulated emission; and hence, sustain laser action. The optical cavity consists of two highly reflective mirrors. One mirror with reflectivity of almost 100% (perfect reflector) completely reflects the

light. The other mirror with a reflectivity of about 80 to 90 % (partial reflector) reflects most part of the light, but allows some part of the light to come out as a laser beam.

Two essential processes to initiate a laser action are (i) stimulated emission and (ii) population inversion. In stimulated emission, the transition occurs when the atom in its excited state interacts with an incident photon (quantum of light energy) at the transition frequency as shown in Figure 7.14. Atoms undergoing stimulated emission radiate photons of the same frequency in the same direction and in phase with the stimulating radiation.

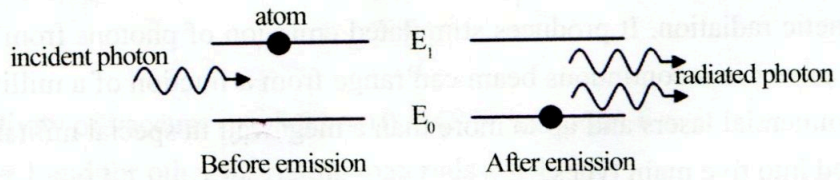


Figure 7.14 Stimulated emission

For stimulated emission to occur, a population inversion would be needed. That means the population of the upper level of a transition would have to be higher than that of the lower level. Then a photon of the transition energy would be more likely to stimulate emission from the excited state than to be absorbed by the lower state. The result is laser gain or amplification, a net increase in the number of photons with the transition energy. To sustain a laser action the lower level must be depopulated while the upper level is populated, otherwise the population inversion will end as shown in Figure 7.15.

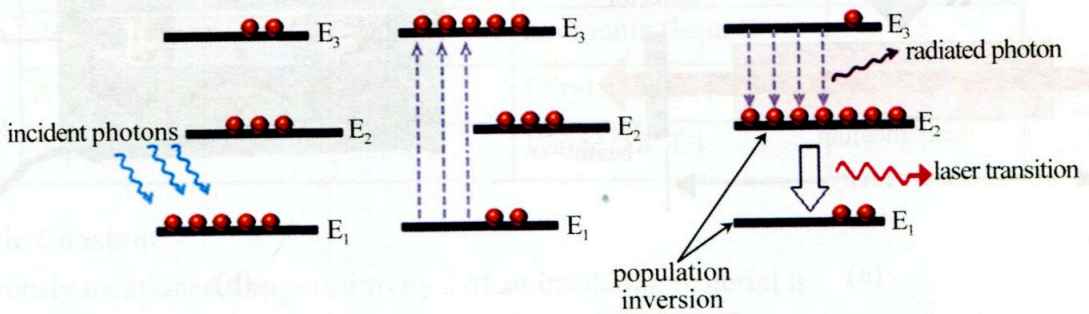


Figure 7.15 Population inversion obtained by pumping with flash lamp

Characteristics of Laser Light

The output lasers beam is narrow, the emission is coherent and the light is monochromatic. A typical beam divergence is around a milliradian, i.e, it will spread only to 1 m in diameter after travelling 1 km. Thus, a modest laser power can produce a high intensity inside the small area of laser beam. For example, 1 mW He-Ne laser beam is much brighter than the sunlight. One of the unique properties of a laser beam is coherence. This means that the light waves it contains are in phase with one another. This allows the wide use of lasers in holography.

Applications of Laser

Lasers are used in various applications; bloodless surgery, destroying kidney stones, eyelens curvature corrections, cancer diagnosis and therapy, fibre-optic endoscopy to detect ulcers in the intestines, in dentistry and cosmetic treatments.

- In communication, laser light is used in optical fibre communications to send information over large distances and also used in space communication such as radars and satellites.
- Laser range finders are used to determine the distance to an object.
- In industries, high power lasers are used to cut metals, drill aerosol nozzles and orifices with the required precision.
- In shops and business establishments, laser light is used to collect the information from the bar code printed on the product.
- Lasers are used in the semiconductor industries for photolithography, in computer printers and to store and retrieve information in compact disc (CD).

Example 7.5 In a laser unit, laser beam of wavelength 6328 Å is emitted. How many photons are released per second if the output power is 1 mW? (Plank constant $h = 6.62 \times 10^{-34}$ J s)

$\lambda = 6328$ Å $= 6328 \times 10^{-10}$ m, $P = 1$ mW$= 10^{-3}$ W

Energy of a photon in laser beam, $E = \dfrac{hc}{\lambda}$

$$= \frac{6.62\times10^{-34}\times3\times10^{8}}{6328\times10^{-10}} = 3.14\times10^{-19} \text{ J}$$

Since output power is 1 mW, the output energy in 1 s is 1 mJ.

$$\text{Number of photons per second} = \frac{\text{Output energy per second}}{\text{Energy of a photon}}$$

$$= \frac{10^{-3}}{3.14\times10^{-19}} = 3.18\times10^{15}$$

Reviewed Exercise

- How many photons are emitted in 1 s by a 7.5 mW of CO_2 laser with the wavelength of 10.6 μm? (Plank constant $h = 6.62 \times 10^{-34}$ J s)

Key Words: population inversion, coherence, pumping, monochromaticity

7.6 FIBRE OPTICS

The fibre optics works on the principle of total internal reflection. Fibre optics refers to the technology that transmits information as light pulses along a glass or plastic fibre. A fibre optic cable can contain a varying number of glass fibres from a few up to hundreds.

An optical fibre is composed of two concentric layers, called the core and cladding which surrounds the glass fibre core. The cladding is surrounded by a layer called sheath (plastic buffer)

which protects the optical fibre as shown in Figure 7.16 (c). The core and cladding have different refractive indices. The refractive index of core will be slightly greater than that of cladding.

If the light ray is injected and strikes the core to cladding interface at an angle greater than the critical angle, it is total internally reflected back into the core as shown in Figure 7.16 (a) and (b). The light ray then continues bouncing down the length of the fibre cable. However, if the angle of incidence at the core to cladding interface is less than the critical angle, both reflection and refraction take place. Because of refraction and absorption, the light beam attenuates and dies off over a certain distance.

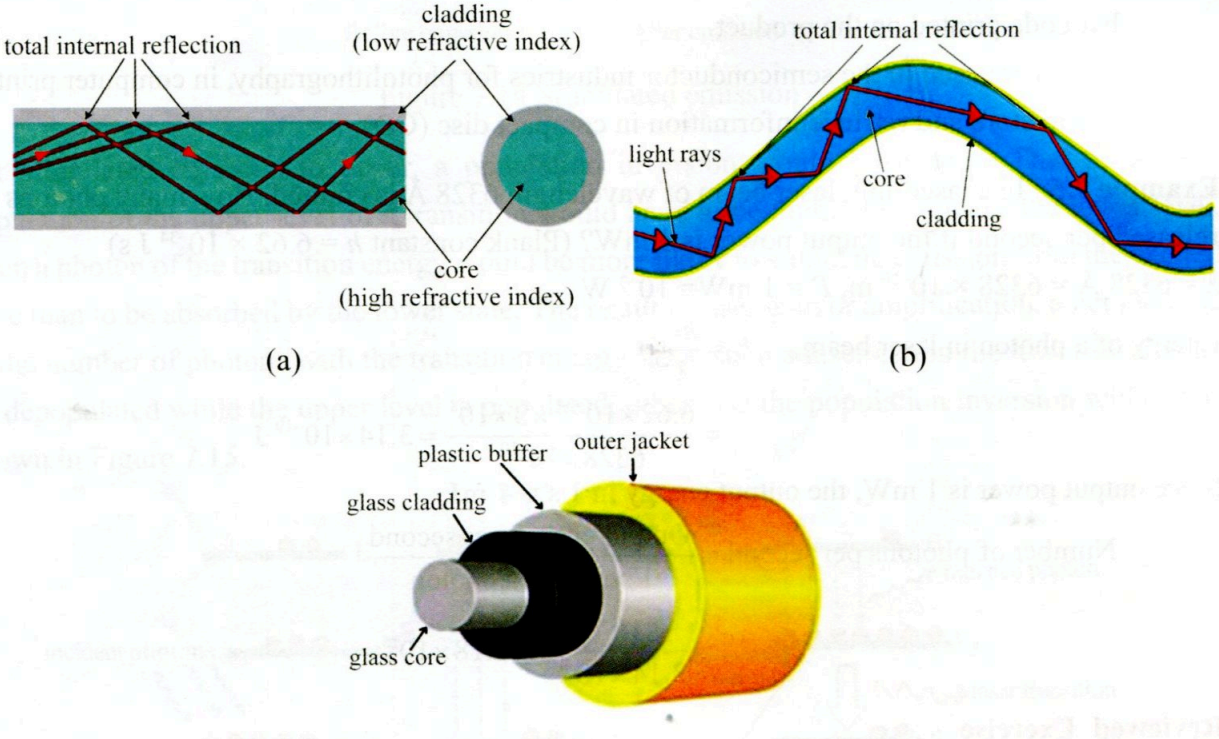


Figure 7.16 (a) Principle of total internal reflection in a section (b) cable of an optical fibre (c) structure of an optical fibre

A useful application of fibre optics is in endoscopy. In medicine, endoscopy is a non-surgical procedure to examine the interior of a hollow organ or cavity of the human body. Endoscopes use optical fibres to observe an image inside the body. A bundle of optical fibres is inserted into the body to clearly monitor the interior organs of the body as shown in Figure 7.17.

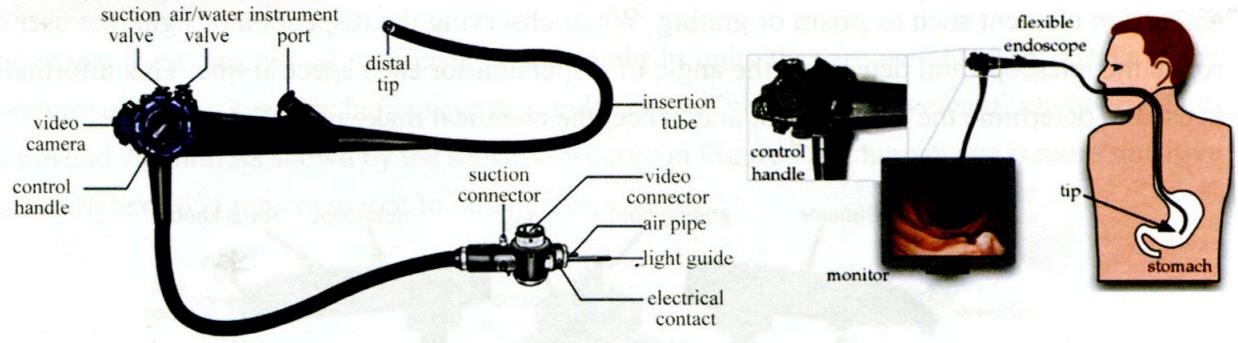


Figure 7.17(a) Endoscope (b) endoscopy of stomach

Example 7.6 The core of an optical fibre has a diameter 2 μm and refractive index of 1.63. The refractive index of the cladding is 1.5. Determine the maximum angle for which the light rays incident on the end of fibre rod to total internal reflection at the core to cladding interface.

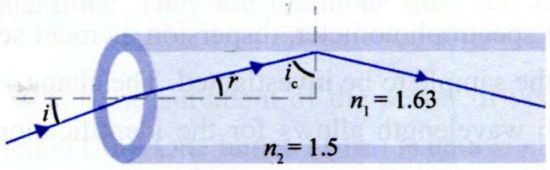
Refractive index of core $n_1 = 1.63$ and that of cladding $n_2 = 1.5$

According to Snell's law, the critical angle for core to cladding interface

$$\sin i_C = \frac{n_2}{n_1}$$

$$= \frac{1.5}{1.63}$$

$$i_C = 67^\circ$$



From the figure, $r = 90^\circ - i_C = 90^\circ - 67^\circ = 23^\circ$

Refraction at air- core interface, $n_1 = \frac{\sin i}{\sin r}$

$$\sin i = 1.63 \times \sin 23^\circ$$

$$i = 39.6^\circ$$

Reviewed Exercise

- What is the function of cladding in an optical fibre?

Key Words: total internal reflection, cladding, endoscopy

7.7 SPECTROMETER

A spectrometer is a scientific instrument used to separate and measure spectral components of a physical phenomenon. It is a widely used scientific tool for many disciplines, including biology, chemistry, agriculture and more.

The optical spectrometer, shown in Figure 7.18, consists of light source, collimator, dispersion

elements and a telescope. The collimated beam from a specific light source is dispersed by dispersion element such as prism or grating. When observing the dispersion of light, the user can rotate the telescope and determine the angle of dispersion for each spectral line. This information is used to determine the wavelength; and hence, the chemical make up of the source.

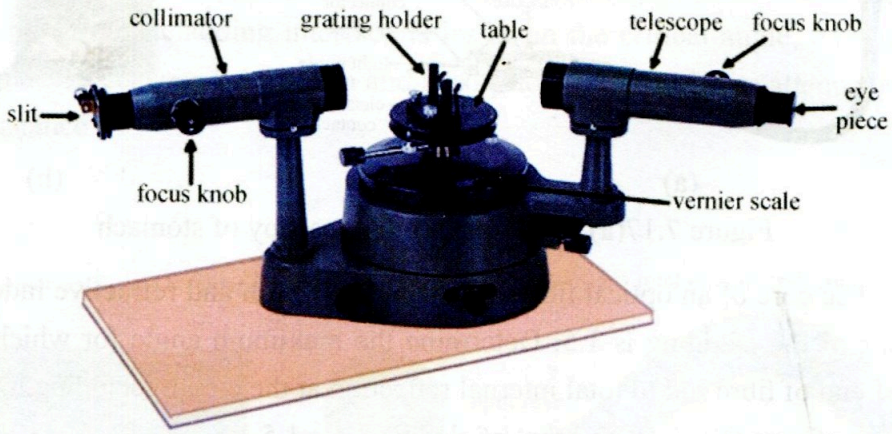


Figure 7.18 Optical spectrometer

In a spectrophotometer, dispersion element separates light into spectrum which is allowed to fall on the sample to be investigated. The change in the absorption and emission of the light intensity with wavelength allows for the identification of the materials as shown in Figure 7.19. Three of the most common optical spectrometers are spectrophotometers (UV-Vis spectrometer), spectrofluorometers and Raman spectrometers.

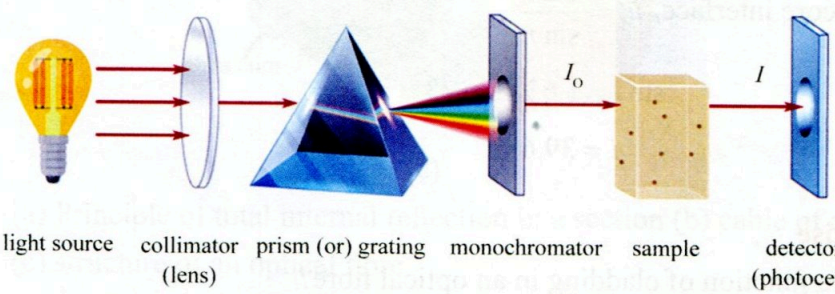


Figure 7.19 Schematics of a spectrophotometer

Reviewed Exercise

1. Name the dispersion elements used in the optical spectrometer.
2. Mention the different types of spectrometer.

Key Words: spectral line, spectrophotometers, spectrofluorometers, Raman spectrometers

7.8 PHOTOMETRY

Photometry is the science of measuring visible light in units that are weighted according to the sensitivity of human eye. The human eye responds only to light having wavelengths between about 400 nm and 700 nm. As shown by the sensitivity curve in Figure 7.20, human eye is more sensitive to green light of 555 nm compared to other colors.

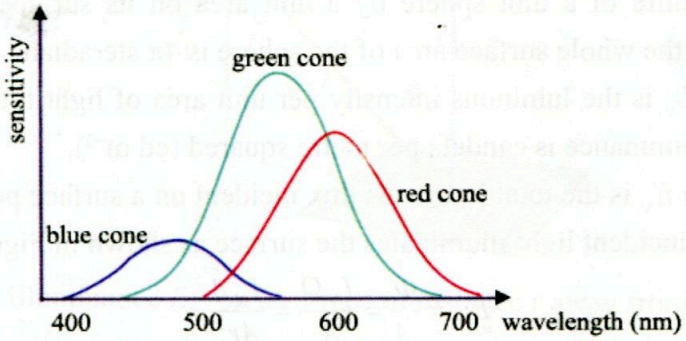


Figure 7. 20 Sensitivity response of human eye depending on wavelength

In photometry, there are four main photometric quantities. They are luminous flux, luminous intensity, luminance, and illuminance.

Luminous flux: Luminous flux (luminous power) Φ is a measurement of the power of visible light produced by a light source. It is measured in lumen (lm). One lumen (1 lm) is equivalent to 1/683 watt of light with a wavelength of 555 nm.

Luminous intensity: Luminous intensity I_V is a measure of the luminous flux emitted by a light source in a direction per unit solid angle shown in Figure 7.21.

$$I_V = \frac{\Phi}{\Omega} \tag{7.7}$$

where I_V is luminous intensity, Φ is luminous flux and Ω is the solid angle.

Luminous intensity is measured in candela (cd) or lumen per steradian. The candela is the SI basic unit.

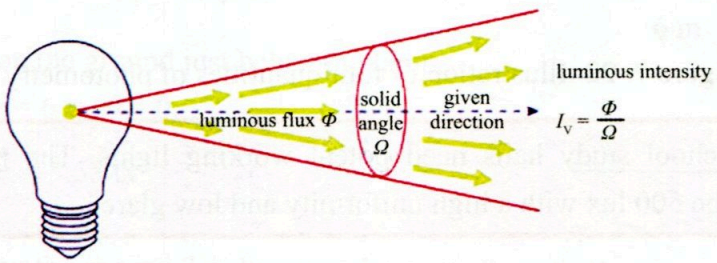


Figure 7.21 Luminous flux and luminous intensity

A solid angle Ω is a measure of the amount of the field of view from some particular point that a given object covers as shown in Figure 7.21.

$$\Omega = \frac{A}{r^2} \qquad (7.8)$$

where Ω is the solid angle, A is the surface area of the spherical cap and r is the radius of the sphere.

In SI system, a solid angle is expressed in steradian (sr). A steradian can be defined as the solid angle subtended at the centre of a unit sphere by a unit area on its surface. The solid angle subtended at the centre by the whole surface area of the sphere is 4π steradian.

Luminance: Luminance L_V is the luminous intensity per unit area of light travelling in a given direction. The SI unit for luminance is candela per metre squared (cd m^{-2}).

Illuminance: Illuminance E_V is the total luminous flux incident on a surface per unit area. It is a measure of how much the incident light illuminates the surface as shown in Figure 7.22.

$$E_V = \frac{\Phi}{A} = \frac{I_V \Omega}{A} = \frac{I_V A}{Ar^2}$$

where A is illuminated surface area, luminous flux of source Φ is I_V Ω , and Ω is solid angle.

$$E_V = \frac{I_V}{r^2} \qquad (7.9)$$

The SI unit for illuminance is lux (lx) (1 lx = 1 lm m^{-2}).

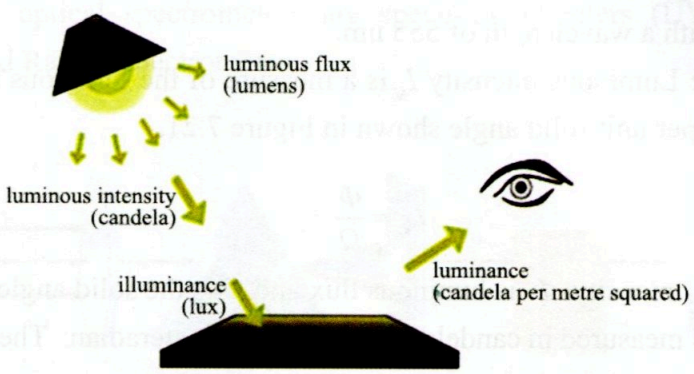


Figure 7. 22 Illustration of four quantities of photometry

Reading areas in school study halls need potent working lights. The minimum average illuminance should be 500 lux with a high uniformity and low glare.

The Inverse Square Law of Illuminance

In Figure 7.23, the illuminance E_V on a surface of distance r away from the source of light of luminous intensity I_V is

$$E_V = \frac{I_V \cos\theta}{r^2} \qquad (7.10)$$

where θ is angle between the direction of light and normal to the illuminated surface.

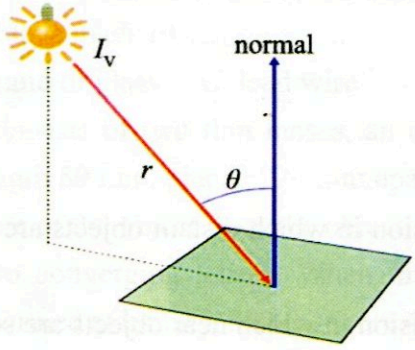


Figure 7.23 Illuminance E_V on a surface of distance r away from the source of light

Therefore, the illuminance E_V at any point on a surface is inversely proportional to the square of the distance r between the source and surface. This is known as inverse square law of illuminance.

Example 7.7 A 10 W fluorescent lamp has a luminous intensity of 35 cd. Find luminous flux it emits.

$I_V = 35$ cd , solid angle for a sphere $= 4\pi\ sr$

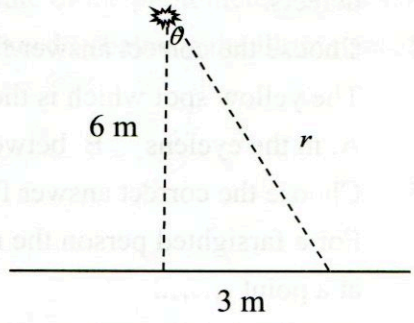
luminous flux emitted from the lamp $\Phi = I_V\,\Omega$

$$= 35 \times 4\pi = 439.8 \text{ lm}$$

Example 7.8 The luminous intensity of a source is 200 cd. The mounting height of the lamp is 6 m from the ground. Find the illuminance on the ground E_V: (i) just below the lamp, (ii) 3 m horizontally away from the lamp on the ground.

luminous intensity of S source $I_V = 200$ cd,

The illuminance on a surface $E_V = \dfrac{I_V \cos\theta}{r^2}$



(i) The illuminance on the ground just below the lamp.

From the figure, $r = 6$ m, $\theta = 0$,

$$E_V = \frac{200 \times \cos 0}{6^2} = 5.55 \text{ lx}$$

(ii) The illuminance on the ground 3 m horizontally away from the lamp.

From the figure, $r = \sqrt{6^2 + 3^2} = 6.708$ m, $\cos\theta = \dfrac{6}{6.708}$

$$E_V = \frac{200}{(6.708)^2} \times \frac{6}{6.708} = 3.98 \text{ lx}$$

Reviewed Exercise

- The luminous intensity of a light source is 500 cd. Find the illuminance of a surface which is 10 m from source if light falls normally on it.

Key Words: photometry, luminous flux, lumen, luminous intensity, candela, illuminance, lux, luminance

SUMMARY

Farsightedness is a defect of vision in which distant objects are seen clearly, but objects nearby are blurry.

Nearsightedness is a defect of vision in which near objects are seen clearly, but objects far away are blurry.

Microscope uses lenses to make enlarged images of objects that are too small for the unaided eye to see.

The optical instrument used for viewing distant objects is called a **telescope**.

A **laser**, acronym for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation, can be considered as a source of a narrow beam of intense, monochromatic, coherent light in the visible, infrared or ultraviolet part of the electromagnetic radiation.

EXERCISES

1. Name the essential parts of a human eye.
2. Discuss the important role of ciliary muscles for the eye.
3. State two defects of vision. Mention of the causes of the defects and corrections to these defects.
4. Choose the correct answer from the following.
   The yellow spot which is the most light-sensitive spot is .....
   A. in the eyelens B. between the eyelens and the retina C. on the retina
5. Choose the correct answer from the following.
   For a farsighted person the rays from the near objects pass through the eyelens and intersect at a point .....
   A. in front of the retina B. on the retina C. behind the retina
6. Draw a ray diagram showing how the rays of light travel in a compound microscope.
7. A concave mirror instead of a lens is used as an objective in an astronomical telescope. Why?
8. How is a terrestrial telescope constructed to obtain the final image which is erect?

9. Choose the correct answer.

   In an optical fibre, the light is transmitted through the .....

   A. core B. cladding C. jacket

10. Choose the correct answer.

    The core of a fibre-optic cable is made of.....

    A. copper wire B. thin strand of glass C. lead wire

11. A compound microscope consists of two thin lenses, an objective of focal length 20 mm and an eyepiece of focal length 50 mm, placed 220 mm apart. If the final image is viewed at infinity, calculate the distance of the object from the objective.

12. A telescope consists of two converging lenses. When in normal adjustment, so that the image of a distant object is formed at infinity, the lenses are 450 mm apart, and the magnifying power is 8. What are the focal lengths of the two lenses? Is the image erect or inverted?

13. A telescope consists of two thin converging lenses of focal lengths 100 cm and 10 cm respectively. It is used to view an object 2000 cm from the objective. What is the separation of the lenses if the final image is 25 cm from the eyelens? Determine the magnifying power for an observer whose eye is close to the eyelens.

14. A He-Ne laser emits photons with energies of 3.14×10^{-19} J. What is the colour of the laser light? (Plank constant $h = 6.62 \times 10^{-34}$ J s)

15. An optical fibre used for communications has a core of refractive index 1.55 which is surrounded by cladding of refractive index 1.45. Calculate the critical angle of the core-cladding boundary.

16. Light from a lamp is falling normally on a surface distant 10 m from the lamp and the illuminance on it is 10 lux. In order to increase the illuminance 9 times, what distance will the lamp be placed from the surface?

17. Two light sources of 8 cd and 12 cd are placed on the same side of the photometer screen at a distance of 40 cm from it. Where should a 80 cd source be placed to balance the illuminance?

CONCEPT MAP

APPLICATIONS OF LIGHT AND OPTICAL INSTRUMENTS

Camera

Essential features

- convex lens
- diaphragm
- aperture
- shutter
- focusing ring
- film

Compound Microscope

Magnifying power

$$MP = \frac{v_o}{u_o} \times \frac{25}{u_e}$$

Electron microscope

Telescope

Astronomical telescope

$$MP = \frac{v_0}{u_e}$$

Terrestrial telescope

Human Eye

Main features

- eyelid
- cornea
- aqueous humour
- vitreous humour
- iris
- pupil
- ciliary muscles
- retina

Common defects of vision

- farsightedness
- nearsightedness

Laser

Fibre optics

Spectrometer

Photometry

- luminous flux
- luminous intensity

$$I_V = \frac{\Phi}{\Omega}$$

- luminance
- illuminance

$$E_V = \frac{I_V}{r^2}$$

CHAPTER 8

INTERFERENCE AND DIFFRACTION OF LIGHT

In this chapter, interference and diffraction of light will be discussed. The study of interference and diffraction, for which the ray approximation of geometrical optics is not valid, are dealt with wave optics (physical optics). Wave optics includes the wave characteristics of light. In wave optics, Huygens' principle explained how the light propagates from a light source.

Learning Outcomes

It is expected that students will

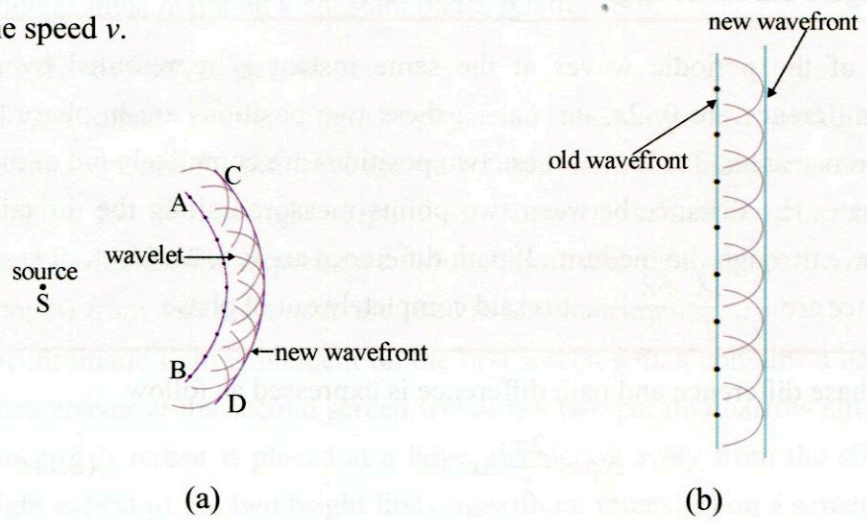
- investigate the Huygens' principle of wave propagation.
- understand the interference of light.
- examine the diffraction of light.
- discuss the pattern obtained from diffraction gratings.
- explain diffraction grating effects.

8.1 HUYGENS' PRINCIPLE

In 1678, the Dutch physicist Christiaan Huygens introduced the wave theory of light. Huygens' principle states that every point on a wavefront may be considered as a source of secondary wavelets. The new position of wavefront is the envelope of the wavelets as shown in Figure 8.1.

If the propagating wave with a frequency f is transmitted through the medium at a speed v, then the secondary wavelets will have the same frequency and speed. These secondary wavelets spread out in all directions with the speed v.

Christiaan Huygens
(1629-1695)



(a) (b)

Figure 8.1 Propagation of (a) spherical wavefront and (b) plane wavefront

Wavefront

A wavefront is defined as a surface associated with a propagating wave which is a locus of all points in the wave that have the same phase. On a particular wavefront, at a given moment of time, all particles of the medium are undergoing the same phase. Wavefront is always perpendicular to the direction of propagation (or) rays. There are two types of wavefront; plane wavefront and spherical wavefront. The spherical wavefront from a source transforms into plane wavefront when it is far away from the source.

Phase (Phase Angle)

The phase angle is the position of a point on a reference circle of a waveform at an instant as shown in Figure 8.2. The angular position of a point of a periodic wave is known as the phase angle. The phase angle is one of the important characteristics of the periodic wave. In Figure 8.2, phase angles of point A and point B are 0 and $\frac{\pi}{2}$ respectively.

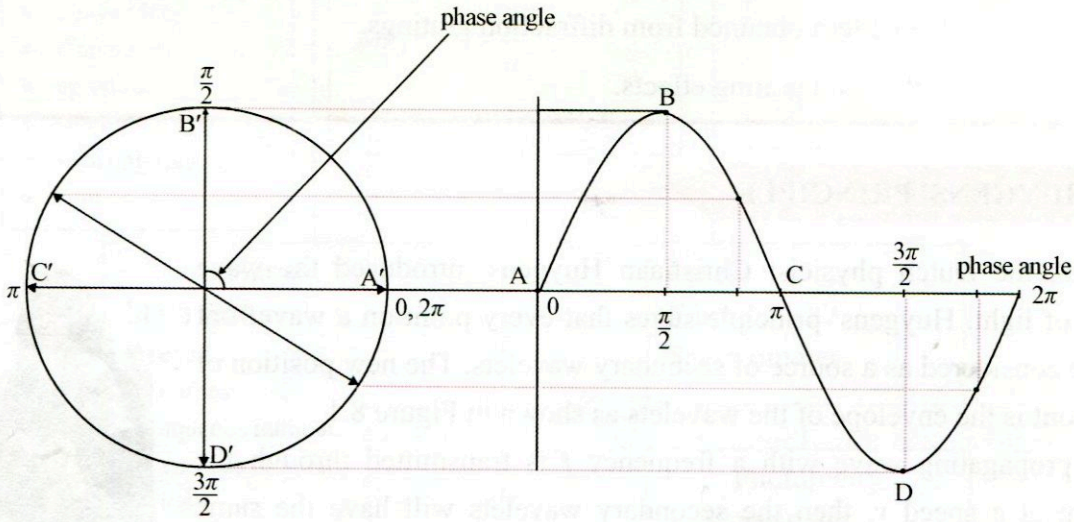


Figure 8.2 Phase angle for a point on a reference circle

The different states of the periodic waves at the same instant is represented by phase difference. If phase difference are 0, 2π, 4π, 6π, ..., these two positions are in phase (same phase). If phase difference are π, 3π, 5π, ..., these two positions are completely out of phase. Path difference indicates the distance between two points measured along the direction of propagation of the wave through the medium. If path difference are 0, λ, 2λ, 3λ, ..., it is said in phase. If path difference are $\frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}$, ..., it is said completely out of phase.

The relation between phase difference and path difference is expressed as follow.

$$\Delta\phi = \frac{2\pi}{\lambda}\Delta x \qquad (8.1)$$

where $\Delta\phi$ is phase difference between two waves of the same wavelength λ and Δx is the path difference of these waves.

Example 8.1 What is the phase difference for two waves of the same frequency arriving at a point with the respective path lengths of 6λ and 4.5λ?

The path difference of two waves $\Delta x = 6\lambda - 4.5\lambda = 1.5\lambda$

The relation between phase difference and path difference is,

$$\Delta\phi = \frac{2\pi}{\lambda}\Delta x$$

$$\Delta\phi = \frac{2\pi}{\lambda} \times 1.5\,\lambda = 3\pi$$

Thus, the two waves are out of phase.

Reviewed Exercise

- How is the phase difference of two waves related to their path difference?

Key Words: wavefront, wavelet, phase angle, wavelength, periodic wave, phase difference, path difference, same phase, out of phase

8.2 INTERFERENCE OF LIGHT

Wave interference is the phenomenon that occurs when two waves meet while traveling along the same medium. Interference of light waves is defined as the modification in the distribution of light energy (intensity) when two or more waves superimpose each other.

If light waves interfere with one another, the maximum and the minimum intensity of light are obtained alternatively. To observe interference of light waves, the two sources of light must be coherent with respect to each other. If the waves emitted from two sources have the same frequency and constant phase difference, the sources are coherent sources. This means that the light sources must maintain a constant phase relationship.

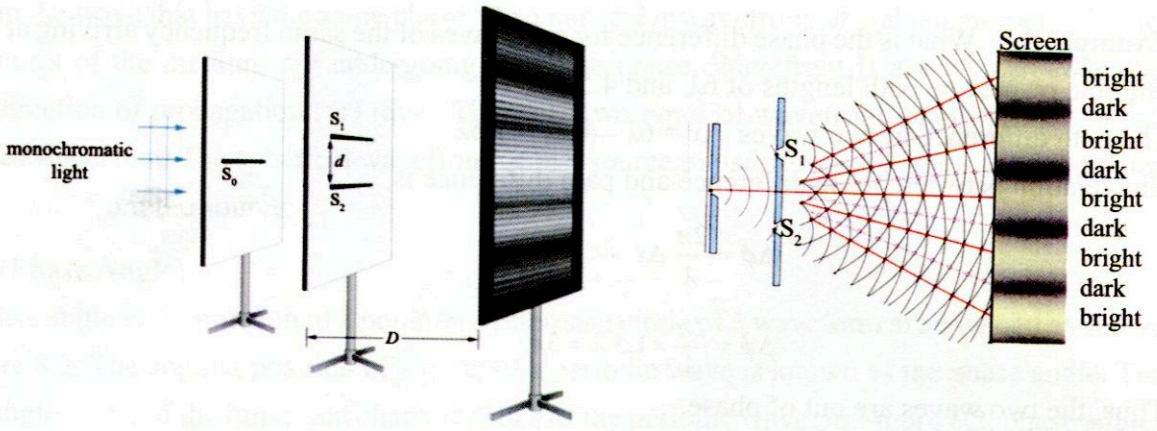
Young's Double-Slit Experiment

In 1801, Thomas Young, English physician and physicist, obtained convincing evidence for the wave nature of light and was able to measure wavelengths for visible light. Figure 8.3 shows a schematic diagram of Young's double-slit experiment. Monochromatic light (a particular wavelength) from a single source are used in his experiment.

Thomas Young
(1773 –1829)

A monochromatic light is incident on the first screen which contains a narrow slit. The emerging light then arrives at the second screen which has two parallel narrow slits S_1 and S_2 separated by a distance d. A screen is placed at a large distance D away from the slits shown in Figure 8.3. We might expect to see two bright lines (maximum intensity) on a screen, but instead a series of

alternate bright and dark lines are seen on the screen. Young was able to explain this result as interference phenomenon.



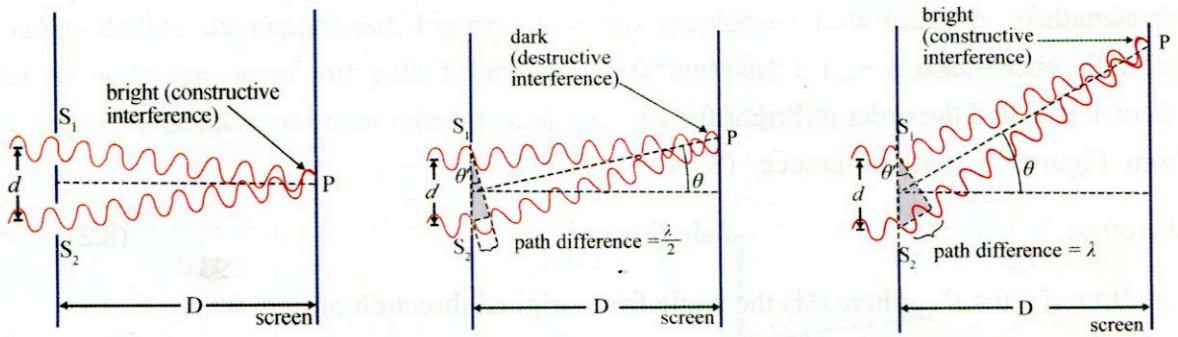
(a) (b)

Figure 8.3 Young's double slits experiment

When a plane wave passes through a slit, a circular wavelet emerges from a slit because the circular wavelet is emitted from a point on the plane wave according to Huygens' principle. When this circular wave passes through two slits, each slit acts as a secondary source of circular waves. At any time, two waves emitted from each slit are of same frequency, same amplitude and are in phase. That is the two waves are coherent. Two waves from each source will interfere and as a result, an interference pattern will be displayed on the screen. The interference pattern consists of a series of bright and dark parallel lines called fringes.

The formation of interference can be explained as follows. The waves spread out in all directions after passing through two slits. Figure 8.4 (a), (b) and (c), show only for three different bending angles θ of waves. In Figure 8.4 (a), the waves reaching the centre of the screen ($\theta = 0$) travel the same distance, so they are in phase. A crest (or trough) of one wave arrives at the same time as a crest (or trough) of the other wave. Hence, constructive interference occurs at the centre of the screen and the bright fringe appears. The bright fringe at centre of the screen is called central bright line.

In Figure 8.4 (b) and (c) light waves arriving the point P from S_1 and S_2 will travel different distances. Light waves from the two sources S_1 and S_2 are initially in phase. When arriving at point P, they may be out of phase or in phase according to the path difference ($S_2P - S_1P$).



(a) (b) (c)

Figure 8.4 (a) At the centre of screen the waves from the two slits travel the same distance and are in phase

(b) the lower wave travels an extra distance of one half wavelength and the waves from two slits are completely out of phase

(c) the lower wave travels an extra distance of one whole wavelength and the waves from two slits are in phase

If the path difference is equal to an integral number of wavelengths ($0, \lambda, 2\lambda, \ldots$) the waves will interfere constructively, leading to a bright fringe on the screen.
If the path difference is equal to a half-integral number of wavelengths ($\frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \ldots$) the waves will interfere destructively, leading to a dark fringe on the screen.

Constructive interference results in bright fringes and destructive interference results in a dark fringe.

Position of Fringes in Young's Double Slit Experiment

The position of particular bright and dark fringes on the screen is given by the distance from the central bright fringe, denoted y as shown in Figure 8.5.

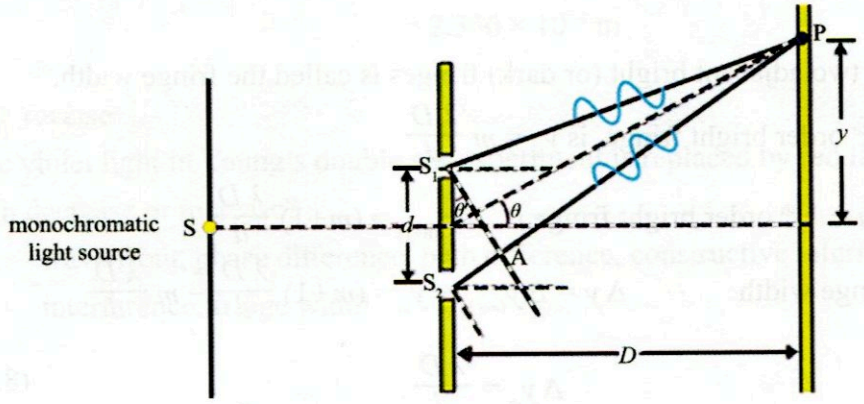


Figure 8.5 Position of fringes

For bright fringe to be formed at P, the path difference is equal to an integral number of wavelengths.

Path difference = $m\lambda$ $(m = 0, 1, 2, 3, \ldots)$

where m is called the order of bright fringe.

From Figure 8.5 path difference $(S_2 A) = d \sin\theta'$

Therefore, $$d \sin\theta' = m\lambda \qquad (8.2)$$

For $D >> d$, $\theta \approx \theta'$, where θ is the angle from original direction of light wave.

Under these conditions θ is small, the approximation can be used θ (rad) $\approx \sin\theta \approx \tan\theta \approx \frac{y}{D}$

Therefore, $d \sin\theta = m\lambda$

$$d\frac{y}{D} = m\lambda$$

$$y = m\frac{\lambda D}{d} \qquad (8.3)$$

where y is distance of the m^{th} bright fringe from the central bright line.

The zeroth order ($m = 0$) bright corresponds to the central bright line at $\theta = 0$, and the first-order brights ($m = 1$) can be seen on either side of the central line.

Similarly, the condition for dark fringe at P, when the path difference is a multiple of half wavelengths, is given by,

For dark fringes, $$d \sin\theta = \left(m + \frac{1}{2}\right)\lambda \qquad (m = 0, 1, 2, 3, \ldots)$$

$$d\frac{y}{D} = \left(m + \frac{1}{2}\right)\lambda$$

$$y = \left(m + \frac{1}{2}\right)\frac{\lambda D}{d} \qquad (8.4)$$

where y is distance of the m^{th} dark fringe from the central bright line.

Fringe width

Distance between two adjacent bright (or dark) fringes is called the fringe width.

The distance of m^{th} order bright fringe is $y_m = m\frac{\lambda D}{d}$

The distance of $(m + 1)^{th}$ order bright fringe is $y_{(m+1)} = (m+1)\frac{\lambda D}{d}$

Fringe width $$\Delta y = \Delta y_{(m+1)} - y_m = (m+1)\frac{\lambda D}{d} - m\frac{\lambda D}{d}$$

$$\Delta y_m = \frac{\lambda D}{d} \qquad (8.5)$$

The position of fringes and fringe width depend on the wavelength of light source.

Intensity Distribution of Fringes

In Young's double slit experiment, Figure 8.6 shows graphically how intensity of interference fringes on the screen varies with path difference. Bright and dark fringes are formed alternatively where bright fringe has maximum intensity and intensity of dark fringe is zero.

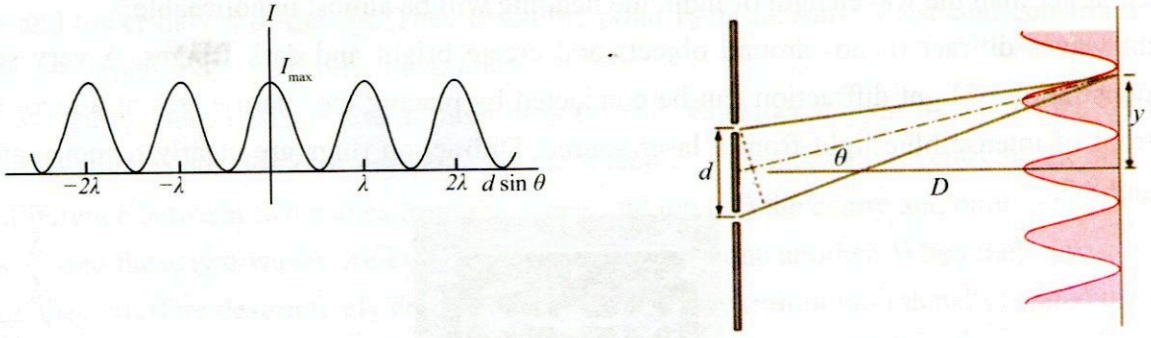


Figure 8.6 Intensity distribution of interference fringes

Example 8.2 In a double slit interference experiment, the distance between the slits is 5 mm and the screen is 2 m from the slits. Yellow light from a sodium lamp is used and it has a wavelength of 5.89×10^{-7} m. Find the distance between the first and second order bright fringes on the screen.

Distance between the slits $d = 5$ mm $= 5 \times 10^{-3}$ m

Distance between the slits and the screen $D = 2$ m

Wavelength $\lambda = 5.89 \times 10^{-7}$ m

The distance between the first and second order bright fringes = fringe width

$$\Delta y = \frac{\lambda D}{d}$$

$$\Delta y = \frac{5.89 \times 10^{-7} \times 2}{5 \times 10^{-3}}$$

$$= 2.356 \times 10^{-4} \text{ m}$$

Reviewed Exercise

- If the violet light in Young's double slit experiment is replaced by red light, will the fringe width decrease or increase?

Key Words: wave front, phase difference, path difference, constructive interference, destructive interference, fringe width

8.3 DIFFRACTION OF LIGHT

Diffraction is defined as the bending of waves around the corners of an obstacle (or) through an aperture into the region of geometrical shadow of the obstacle (or) aperture. The amount of bending depends on the wavelength of light relative to the size of the aperture. If the aperture is much larger than the wavelength of light, the bending will be almost unnoticeable.

Light waves diffract (bend) around objects and create bright and dark fringes. A very simple demonstration of light diffraction can be conducted by placing the sharp edges of a razor blade in front of intense blue light from a laser source. Diffraction rings are clearly demonstrated in Figure 8.7.

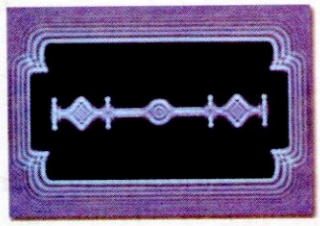


Figure 8.7 Diffraction of light at the edges of a razor blade

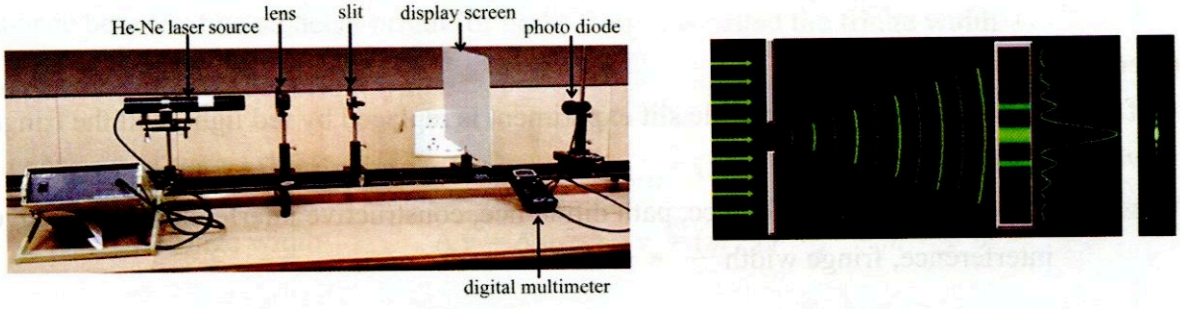
There are two main classes of diffraction, which are known as Fresnel diffraction and Fraunhofer diffraction.

Fresnel diffraction: The source of light and the screen are at a finite distance from the obstacle causing the diffraction.

Fraunhofer diffraction: There is an infinite distance between the source and the screen from the obstacle. In this section Fraunhofer diffraction will be discussed.

Diffraction by a Single Slit (Fraunhofer diffraction)

The experimental arrangement for Fraunhofer diffraction is shown in Figure 8.8. When light passes through a single slit whose width a is of the order of the wavelength of the light, then the diffraction pattern is focussed on a screen that is a distance $D >> a$ away from the slit.



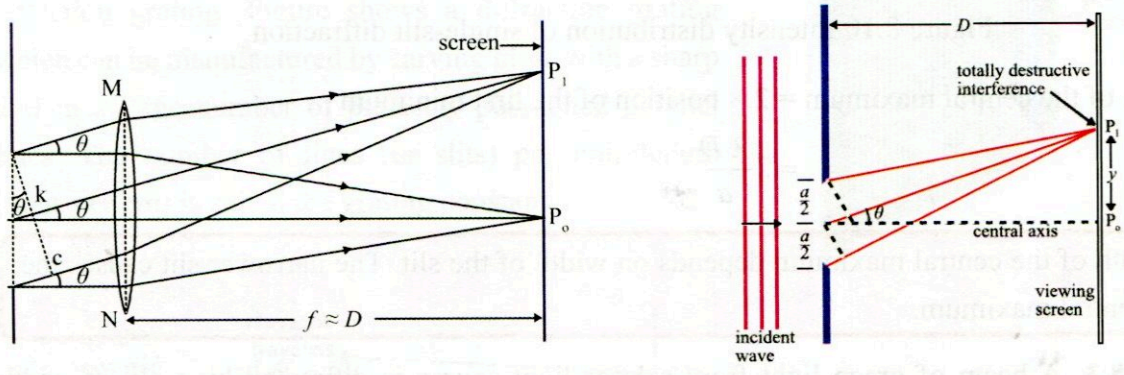
(a) (b)

Figure 8.8 The experimental arrangement for a single slit diffraction

The diffraction pattern consists of a bright central portion (the central maximum). On either side there are alternating bright and dark fringes shown in Figure 8.8 (b). These bright and dark fringes are created by interference of diffracted light waves. Each successive bright fringes becomes less intense proceeding outward, away from the central maximum.

Consider all the secondary wavelets that can be sent out from any two corresponding points of the upper and lower halves of the slit. They reach the point P_o in the same phase and constructively interfere to produce a central bright maximum.

Now we will explain how the first minimum is formed. Consider the slit to be divided by two parts. The angular position θ of any point on the screen is measured from centre of the slit. The path difference between two waves from the centre and top (or) the centre and bottom edge of the slit is $\frac{\lambda}{2}$ and these two waves are exactly out of phase with one another. When they arrive at the screen, they interfere destructively and produces a dark fringe (minimum intensity) called the first minimum. The first minima are seen on the either side central bright maximum.



(a) (b)

Figure 8.9 (a) Central bright maximum at P_o
(b) Path difference for two waves from the centre and top (or) from the centre and bottom

In Figure 8.9 (b), to get the first minimum (dark) the path difference between two waves from the centre and top (or) the centre and bottom edge of the slit is $\frac{a}{2}\sin\theta$.

Thus, for the first minimum $\frac{a}{2}\sin\theta = \frac{\lambda}{2}$

$$a \sin\theta = \lambda$$

For the second minimum, the slit is divided into four parts. The path difference between two waves from the corresponding points from each quarter of the slit is $\frac{a}{4}\sin\theta$.

Thus, for the second minimum $\frac{a}{4}\sin\theta = \frac{\lambda}{2}$

$$a \sin\theta = 2\lambda$$

Generally for the m^{th} minimum, $a \sin\theta = m\lambda$ where $m = 1, 2, 3, \ldots$ (8.6)

The position of m^{th} minimum y is the distance from the central maximum.

From Figure 8.9, $y = D\tan\theta \approx D\sin\theta \approx m\dfrac{\lambda D}{a}$

The intensity distribution for single-slit diffraction is shown in Figure 8.10.

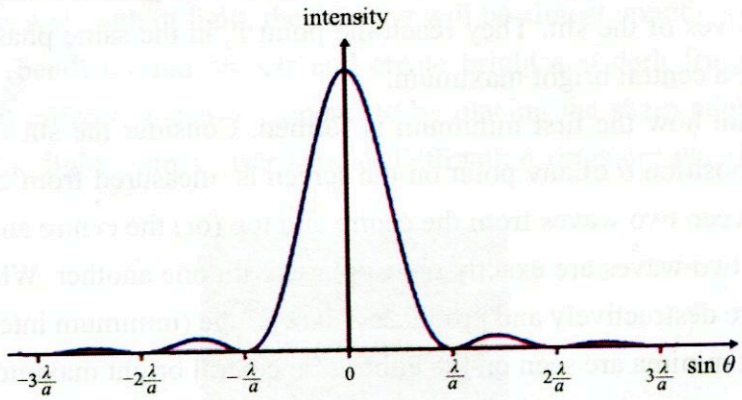


Figure 8.10 Intensity distribution of single-slit diffraction

The width of the central maximum = 2 × position of the first minimum

$$= 2\frac{\lambda D}{a}$$

The width of the central maximum depends on width of the slit. The narrower slit causes the wider central maximum.

Example 8.3 A beam of green light from a laser light source is diffracted by a slit of width 0.55 mm. The diffraction pattern forms on a wall 2.06 m beyond the slit. The distance between the positions of the first minima on both sides of central bright fringe is 4.1 mm. Calculate the wavelength of the laser light.

Width of the slits $a = 0.55$ mm $= 0.55 \times 10^{-3}$ m

Distance between the slits and the wall $D = 2.06$ m

The width of the central maximum, $2\dfrac{\lambda D}{a} = 4.1$ mm

$$\lambda = \frac{4.1\times10^{-3} a}{2D}$$

$$\lambda = \frac{4.1\times10^{-3}\times0.55\times10^{-3}}{2\times2.06}$$

$$= 0.5473\times10^{-6}\ \text{m} = 547.3\ \text{nm}$$

Reviewed Exercise

1. When the slit width in a single slit experiment is reduced, what will be affect on the diffraction pattern?

2. What is the difference between Fresnel diffraction and Fraunhofer diffraction?

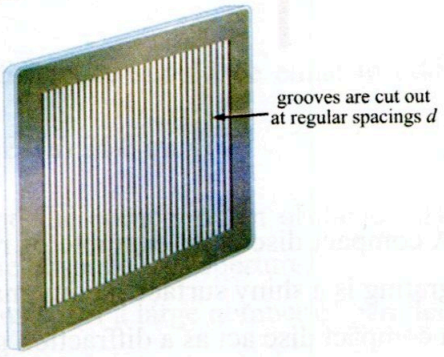
Key Words: Fraunhofer diffraction, diffraction pattern, central maximum, minimum intensity

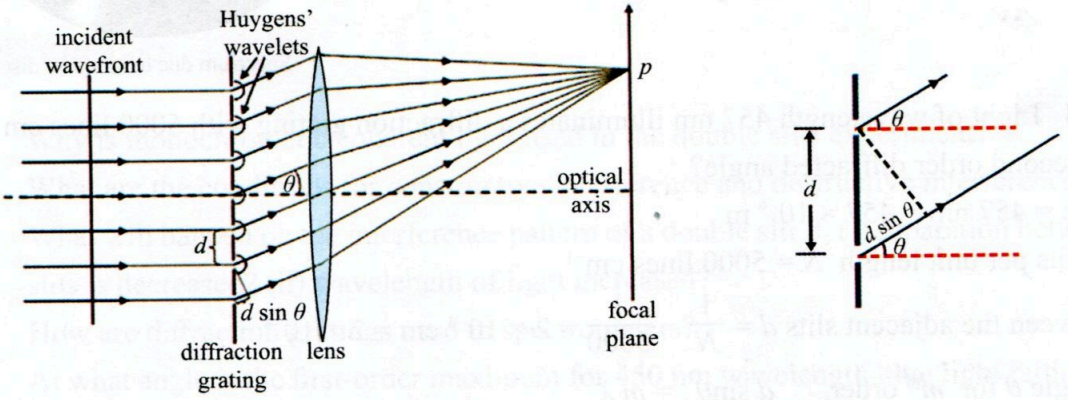
8.4 DIFFRACTION GRATING

Most modern-day applications of slit interference use not just two slits but many, approaching several thousand per centimetre for practical purposes. This optical instrument is called the diffraction grating, an important tool in optical analysis.

A diffraction grating is an optical device which consists of a large number of parallel slits of same width separated by equal opaque spaces. A diffraction grating can disperse light into its wavelength components. Therefore, it is used for the purpose of resolving light into spectra. Diffraction grating is thus widely used as dispersive elements in spectrographic instruments and a crucial tool in the study of atomic and molecular compositions.

Two types of grating are transmission grating and reflection grating. Figure shows a diffraction grating which can be manufactured by carving glass with a sharp tool in a large number of precisely positioned parallel lines. The number of lines (or slits) per unit length (lines per cm) is called the grating constant.





(a) (b)

Figure 8.11 Diffraction of light in each slit of diffraction grating

The condition for maximum intensity is the same as that for a Young's double slit.

For maximum intensity to be formed at P, the path difference of two waves from adjacent slits is equal to an integral number of wavelengths.

Path difference = $m\lambda$ ($m = 0, 1, 2, 3, \ldots$)

From Figure 8.11 Path difference = $d \sin\theta$

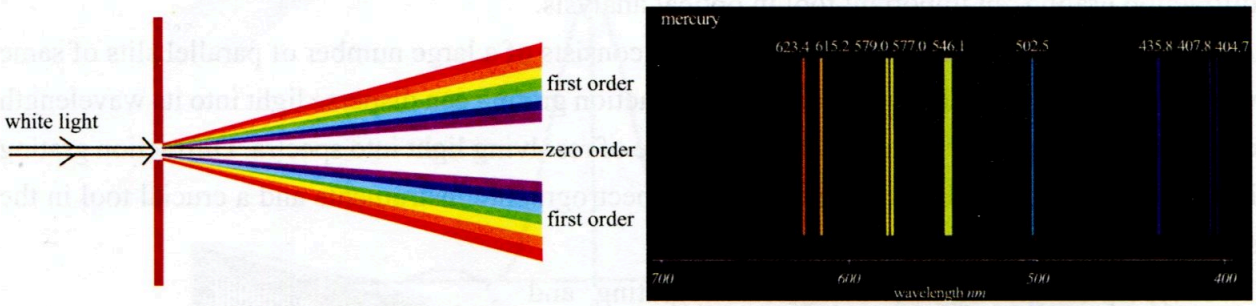
Therefore, $d \sin\theta = m\lambda$

where d is the distance between the adjacent slits, θ is the angle of diffraction, λ is the wavelength of the light and m is the order of maxima.

If the number of slits per unit length is N, d is equal to $\frac{1}{N}$.

In Figure 8.12, there is no dispersion of incident white light at the centre (i.e, $m = 0$).

In the first order diffraction spectrum, longer wavelength (red light) diffracts at a greater angle than shorter wavelength (violet light).



(a) (b)

Figure 8.12 (a) Spectrum for white light (b) mercury spectrum

A compact disc is an example of reflection gratings. A reflection grating is a shiny surface having many fine grooves. The tracks of a compact disc act as a diffraction grating, producing a separation of the colors of white light shown in Figure.

Spectrum due to compact disc

Example 8.4 Light of wavelength 457 nm illuminates a diffraction grating with 5000 lines cm^{-1}. What is the second order diffracted angle?

Wavelength $\lambda = 457$ nm $= 457 \times 10^{-9}$ m

Number of slits per unit length $N = 5000$ lines cm^{-1}

Distance between the adjacent slits $d = \frac{1}{N} = \frac{1}{5000} = 2 \times 10^{-4}$ cm $= 2 \times 10^{-6}$ m

Diffracted angle θ for m^{th} order, $d \sin\theta = m\lambda$

for second order $m = 2$

$2 \times 10^{-6} \sin\theta = 2 \times 457 \times 10^{-9}$

$\theta = 27.19^{o}$

Reviewed Exercise

1. Why can a grating with smaller grating constant produce the wider the diffraction pattern?
2. Why does the compact disc show the diffraction pattern?
3. What are the differences between the interference and diffraction?
4. What is the difference between the diffraction grating and prism?

Key Words: diffraction grating, spectrographic instruments, grating constant, spectrum

SUMMARY

A wave front is defined as a surface associated with a propagating wave which is a locus of all points having the same phase in the wave.

The angular position of a periodic wave is known as the **phase angle**.

If the waves emitted from two sources have the same frequency and constant phase difference, the sources are **coherent sources**.

Interference of light waves is defined as the modification in the distribution of light energy (intensity) when two or more waves superimpose each other.

The condition for **constructive interference** is that the path difference should be equal to an integral multiple of the wavelength.

The condition for **destructive interference** is that the path difference should be equal to odd integral multiple of half wavelength.

Monochromatic light means that it has a single wavelength.

Diffraction of light is defined as the bending of waves around the corners of an obstacle (or) through an aperture into the region of geometrical shadow of the obstacle (or) aperture.

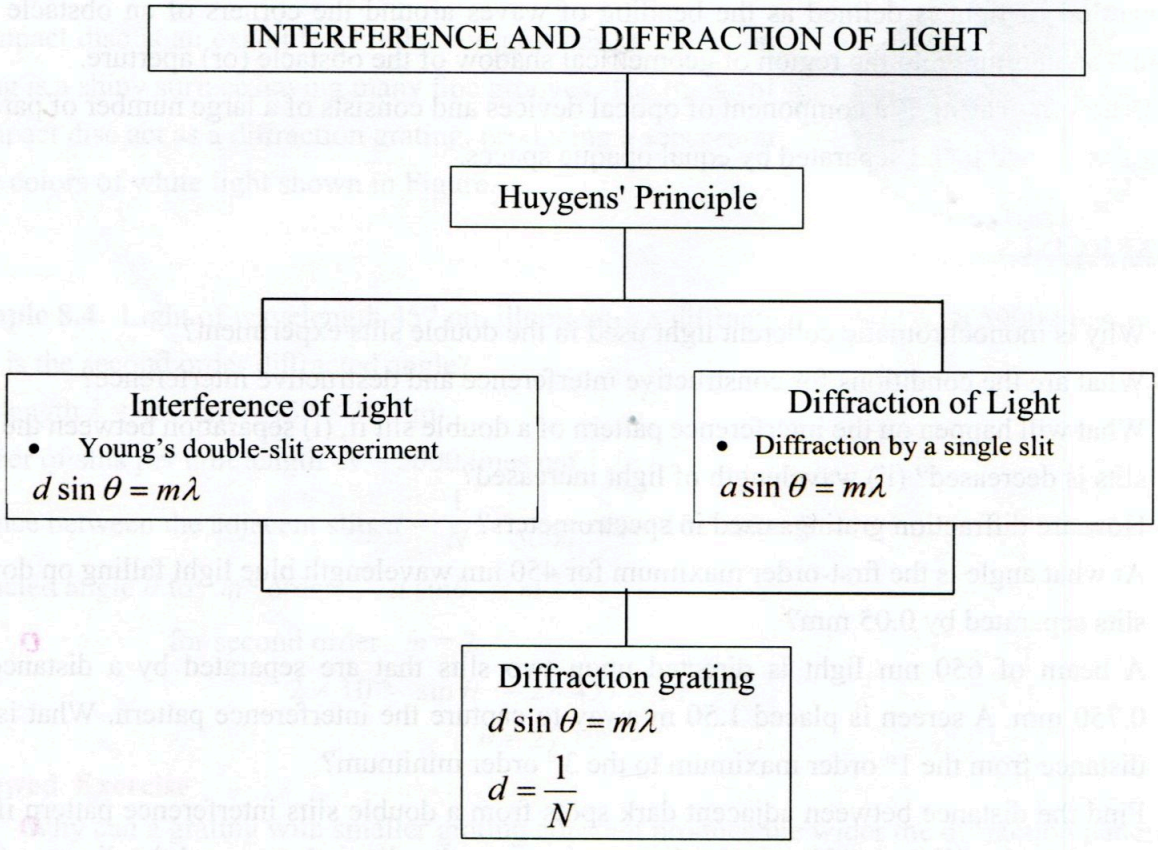
A diffraction grating is a component of optical devices and consists of a large number of parallel slits of same width and separated by equal opaque spaces.

EXERCISES

1. Why is monochromatic coherent light used in the double slits experiment?
2. What are the conditions for constructive interference and destructive interference?
3. What will happen on the interference pattern of a double slit if, (i) separation between the two slits is decreased? (ii) wavelength of light increased?
4. How are diffraction gratings used in spectrometers?
5. At what angle is the first-order maximum for 450 nm wavelength blue light falling on double slits separated by 0.05 mm?
6. A beam of 650 nm light is directed upon two slits that are separated by a distance of 0.750 mm. A screen is placed 1.50 m away to capture the interference pattern. What is the distance from the 1st order maximum to the 3rd order minimum?
7. Find the distance between adjacent dark spots from a double slits interference pattern if the wavelength of light is 500 nm, the distance between the slits is 1 mm, and the distance from the slit to the screen is 2 m.

8. Light of wavelength 580 nm is incident on a slit of width 0.3 mm. The observing screen is placed 2 m from the slit. Find the positions of the first dark fringes and the width of the central bright fringes.
9. If the distance between two slits is 0.05 mm and the distance to a screen is 2.5 m, find the spacing between the first- and second-order bright fringes for yellow light of 580 nm wavelength.
10. A grating containing 4000 slits per centimetre is illuminated with a monochromatic light and produces the second-order bright line at a 30° angle. What is the wavelength of the light used?
11. A diffraction grating has 2000 lines per centimetre. At what angle will the first-order maximum be for 520 nm wavelength green light?
12. How many lines per centimetre are there on a diffraction grating that gives a first-order maximum for 470 nm blue light at an angle of 25°?

CONCEPT MAP



CHAPTER 9

CAPACITOR AND CAPACITANCE

A capacitor is used as an electric circuit component to store electrical energy. The capacitor is similar to a rechargeable battery. The capacitor was originally known as a condenser. In this chapter, we shall study the capacitors, capacitance and applications of capacitors.

Learning Outcomes

It is expected that students will

- examine various types of capacitors and their applications.
- investigate the capacitance and dielectric constant.
- understand the energy stored in a capacitor.
- explain the charging and discharging of the capacitors.
- deduce the combination of capacitors.

9.1 CAPACITORS

A capacitor is an electrical device that stores electrical energy in the form of an electric field. A capacitor consists of two conductors separated by a small distance. An insulator is inserted between its conductors.

Types and Applications of Capacitors

The various types of capacitors and their applications are described below.

Film Capacitors: Film capacitors are the ones that use plastic film as the dielectric medium. They are available in nearly any value and working voltages up to 1500 V. They range from 10 % to 0.01 % in tolerance.

Ceramic Capacitors: Ceramic capacitors are the ones that use ceramic as the dielectric material. It is used in high-frequency circuits such as audio frequency (*af*) to radio frequency (*rf*). In ceramic capacitors, one can develop both high capacitance and low capacitance by altering the thickness of the ceramic disc. Ceramic capacitors are mainly used for high stability performances with low losses. The capacitance values of these capacitors are stable with respect to the applied voltage, frequency and temperature.

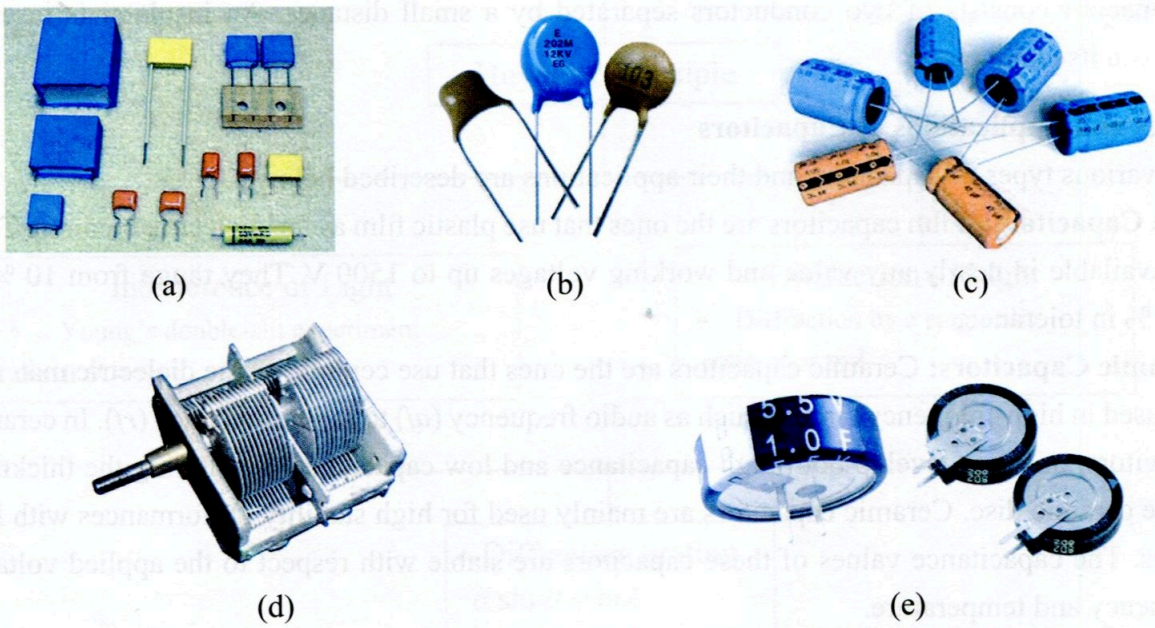
Electrolytic Capacitors: Electrolytic capacitors are the ones that use the oxide layer as the dielectric material. It has a wide tolerance capacity. There are mainly two types of electrolytic capacitors; tantalum and aluminum. They are available with working voltages of up to approximately 500 V. The failure of electrolytic capacitors can be hazardous, resulting in an explosion or fire. Electrolytic capacitors are used for filtering and smoothing rectified alternating

voltage, buffering and interim storage for direct voltage supplies, uninterruptible power supplies and energy storage in flash devices.

Variable Capacitors: Variable capacitors mostly use air as the dielectric medium. The capacitance can be mechanically adjusted. For example, this form of the capacitor is used to set the resonance frequency in tuner devices. Variable capacitors are mostly used in AC circuits especially for high frequency, high power and low loss. They are used in radio tuning applications, antenna tuning and so on.

Supercapacitors: A supercapacitor (SC), also called an ultracapacitor, is a high-capacity capacitor with a capacitance value much higher than other capacitors, but with lower voltage limits, that bridges the gap between electrolytic capacitors and rechargeable batteries. It typically stores 10 to 100 times more energy than electrolytic capacitors. It can accept and deliver charge much faster than batteries, and tolerates many more charge and discharge cycles than rechargeable batteries.

Supercapacitors are used in applications requiring many rapid charge/discharge cycles, rather than long-term compact energy storage in automobiles, buses, trains, cranes and elevators, where they are used for regenerative braking and short-term energy storage or burst-mode power delivery. Smaller units are used as power backup for static random-access memory (SRAM).



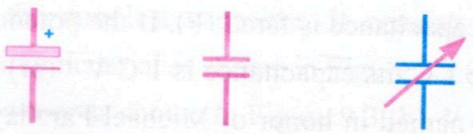
(a) (b) (c)

(d) (e)

Figure 9.1 The various types of capacitors (a) film capacitors (b) ceramic capacitors (c) electrolytic capacitors (d) variable capacitor (e) supercapacitors

Film capacitors and ceramic capacitors are non-polarized capacitors which means that they do not have polarities (positive and negative terminals).

The symbols of a polarized capacitor, a fixed capacitor and a variable capacitor are shown in Figure 9.2.



(a) (b) (c)

Figure 9.2 Symbols of (a) polarized capacitor (b) fixed capacitor (c) variable capacitor

The capacitors are used in both electrical and electronic applications such as filters, energy storage systems, engine starters, signal processing devices, etc.

Capacitors are used;

- for storing energy, which can be used by the device for temporary power outages whenever they need additional power.
- for blocking DC current after getting fully charged but allow the AC current to pass through it.
- as filters in power supplies and energy-storing devices in electronic flash units. They are also used to tune the frequency of radio receivers and to eliminate sparking in automobile ignition systems.
- as the sensor for measuring humidity, fuel levels, mechanical strain, etc.

Capacitors can also be used in a time-dependent circuit (timer circuit). This could be connected to any LED or loudspeaker system, and it is likely that any flashing light or regular beeping uses a timing capacitor.

Reviewed Exercise

- Which capacitor is used as tuning capacitor?

Key Words: film capacitor, ceramic capacitor, electrolytic capacitor, varible capacitor

9.2 CAPACITANCE OF A CAPACITOR

When a capacitor is connected to a battery (an electrical power source) the two conductors of the capacitor have charges of equal magnitude and opposite signs. If one conductor has a charge $+Q$ the other has a charge $-Q$. The magnitude of the charge on each conductor, Q, is called the charge of the capacitor. The potential difference between two conductors of the capacitor, V, is called the potential difference of the capacitor. The quantity of charge Q on the capacitor is linearly proportional to the potential difference V between the conductors; that is, $Q \propto V$. This relationship can be written as $Q = C\,V$, where C is called capacitance of the capacitor.

The capacitance of a capacitor is defined as the ratio of the charge to the potential difference of the capacitor.

The capacitance is represented as,

$$C = \frac{Q}{V} \qquad (9.1)$$

where C is capacitance of the capacitor, Q is charge of the capacitor and V is potential difference between two conductors of the capacitor.

In the SI system, the unit of capacitance is farad (F). If the potential difference of the capacitor is 1 V when it is given a charge 1 C, its capacitance is 1 C V^{-1} (or) 1 F. Therefore 1 F is equivalent to 1 C V^{-1}. The unit farad is named in honor of Michael Faraday. The farad is a very large unit of capacitance. In practice, typical devices have capacitances ranging from microfarad (μF) to picofarad (pF). The sub-multiple units of farad are

$$\begin{aligned} 1 \text{ microfarad } (\mu F) &= 10^{-6} \text{ F} \\ 1 \text{ nanofarad } (nF) &= 10^{-9} \text{ F} \\ 1 \text{ picofarad } (pF) &= 10^{-12} \text{ F} \end{aligned}$$

When the charge on the capacitor is increased, the potential difference also increases proportionally, as mentioned previously. Therefore, the capacitance of a capacitor remains constant.

Although the capacitance of a capacitor does not depend on Q and V, it depends on the size and the shape of the capacitor, and on the nature of the insulator between the two conductors.

Like a capacitor, if a conductor is given some charge its potential will also change. The amount of charge given to a conductor to change its potential by one unit is called the electric capacity of the conductor. The electric capacity is also expressed by Eq. (9.1) and its SI unit is also farad.

Example 9.1 A capacitor has a capacitance of 5 μF. How much of the charge should be removed in order that the potential difference between its plates decreases by 40 V?

Capacitance of capacitor $C = 5$ μF,

Change in potential difference $\Delta V = 40$ V,

Let the charge removed be ΔQ,

$$C = \frac{\Delta Q}{\Delta V}$$

$$\Delta Q = C \Delta V$$

$$= 5 \times 40 = 200 \text{ μC}$$

Hence, 200 μC charge should be removed.

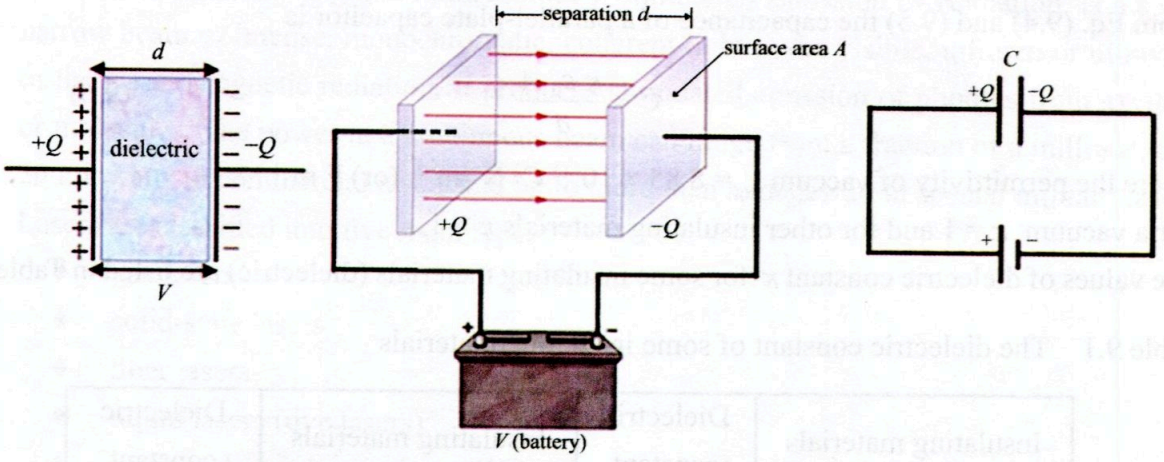
Reviewed Exercise

1. When the charge on a capacitor is increased, does its capacitance increase? Explain.
2. A capacitor stores charge Q at a potential difference V. What happens if the voltage applied to the capacitor is doubled?

Key Words: capacitor, capacitance, electric capacity, charge of capacitor, potential difference, farad

9.3 PARALLEL-PLATE CAPACITOR

A parallel-plate capacitor is the simplest capacitor. It consists of two parallel metal plates separated by air or other insulating material as shown in Figure 9.3 (a). Charging a capacitor means connecting the capacitor to a battery as shown in Figure 9.3 (b). It can be presented by an electric circuit diagram as shown in Figure 9.3 (c).



(a) (b) (c)

Figure 9.3 (a) Parallel-plate capacitor (b) charging a capacitor (c) circuit diagram

Two parallel, metallic plates of equal surface area A are separated by a distance d as shown in Figure 9.3 (a) and (b). One plate carries a charge $+Q$, and the other carries a charge $-Q$. The surface charge density (charge per unit surface area) on each plate is $\sigma = Q/A$. If the plates are very close together, the electric field between the plates is uniform.

Suppose that an insulating material of permittivity ε is placed between the plates. The value of the electric field between the plates can be expressed in terms of surface charge density σ as

$$E = \frac{\sigma}{\varepsilon}$$

$$E = \frac{Q}{\varepsilon A} \qquad (9.2)$$

Because the electric field between the plates is uniform, the magnitude of the potential difference V between the plates equals $E\,d$, therefore,

$$V = E\,d = \frac{Q\,d}{\varepsilon A} \qquad (9.3)$$

The capacitance of the parallel plate capacitor is

$$C = \frac{Q}{V} = \frac{Q}{Q\,d/\varepsilon A}$$

$$C = \frac{\varepsilon A}{d} \qquad (9.4)$$

That is, for a given insulating material the capacitance of a parallel-plate capacitor is directly proportional to the area of its plates and inversely proportional to the separation of the plates.

The permittivity ε of an insulating material is related to that of a vacuum ε_0 as

$$\varepsilon = \kappa \varepsilon_0 \tag{9.5}$$

Here κ is the dielectric constant or the relative permittivity of the insulating material.

From Eq. (9.4) and (9.5) the capacitance of a parallel-plate capacitor is

$$C = \frac{\kappa \varepsilon_0 A}{d} \tag{9.6}$$

where the permittivity of vaccum $\varepsilon_0 = 8.85 \times 10^{-12}\ C^2\ N^{-1}m^{-2}$ (or) $F\ m^{-1}$

For a vacuum $\kappa = 1$ and for other insulating materials $\kappa > 1$.

The values of dielectric constant κ for some insulating materials (dielectric) are listed in Table 9.1.

Table 9.1 The dielectric constant of some insulating materials

Insulating materials	Dielectric constant	Insulating materials	Dielectric constant
Vacuum	1	Nylon (solid)	3.8
Air (1 atm)	1.0006	Mica	3 – 6
Waxed paper	2	Glass	5 – 8
Plywood	2.1	Marble	6
Rubber (hard)	3	Ammonia (liquid)	25
Amber	3	Ethyl alcohol (0 °C)	28.4
Paper	3.6	Water (18 °C)	81

Dielectric Constant

As previously mentioned, the permittivity ε of an insulating material is

$$\varepsilon = \kappa \varepsilon_0$$

$$\kappa = \frac{\varepsilon}{\varepsilon_0} \tag{9.7}$$

Thus, the dielectric constant of an insulating material is the ratio of its permittivity to that of vacuum.

If a parallel-plate capacitor has an insulating material between its plates, according to Eq. (9.6), its capacitance C is

$$C = \frac{\kappa \varepsilon_0 A}{d}$$

If there is a vacuum between its plates, its capacitance C_0 is

$$C_0 = \frac{\varepsilon_0 A}{d}$$

Hence, the dielectric constant κ of the insulating material is

$$\kappa = \frac{C}{C_0} \tag{9.8}$$

Therefore, the dielectric constant can be defined as the ratio of two capacitances as follows.

The ratio of the capacitance of a capacitor with an insulating material between its two conductors to the capacitance of that capacitor with a vacuum between its two conductors is called the dielectric constant of that insulating material.

When a charged capacitor is disconnected from a battery and the air medium between its plates is replaced by an insulating material (dielectrics) as shown in Figure 9.4, the potential difference of the capacitor is found to decrease. Since the charge Q of the capacitor does not change, it indicates that the capacitance of the capacitor increases.

With vacuum $C_0 = \frac{Q}{V_0}$ and with insulating material $C = \frac{Q}{V}$.

Since $V < V_0$, $C > C_0$

$$\frac{C}{C_0} = \frac{V_0}{V} = \kappa \tag{9.9}$$

According to Eq. (9.9), dielectric constant can be expressed as the ratio of capacitances as well as the ratio of potential differences.

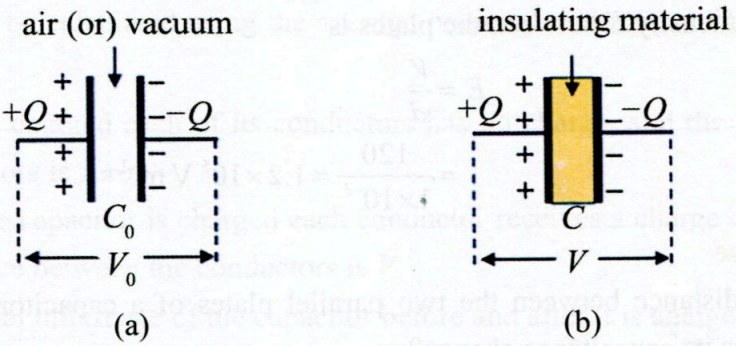


Figure 9.4 A charged capacitor (a) before and (b) after insertion of an insulating material (dielectrics) between the plates

Insulating materials have values of dielectric constant κ greater than unity; that is, their dielectric strength is greater than that of air. Therefore, a dielectric provides the following advantages:

- increase in capacitance,
- increase in maximum operating voltage,
- possible mechanical support between the plates, which allows the plates to be close together without electrical contact, thereby decreasing d and increasing C.

A dielectric increases the maximum operating voltage of a capacitor even though the physical size of the capacitor doesn't change. The dielectric may be able to withstand a larger electric field (higher potential difference) between the plates than air.

Example 9.2 The area of each plate of a parallel-plate capacitor is 1 m^2 and the distance between two plates is 1 mm. If the potential difference between the plates is 120 V and the dielectric constant of the material inserted between them is 3, find (i) the capacitance of the parallel-plate capacitor, (ii) the magnitude of the charge on each plate, (iii) the electric field intensity between the plates.

(i) Area of each plate $A = 1\ m^2$, distance between two plates $d = 1\ mm = 1 \times 10^{-3}\ m$, dielectric constant of the material $\kappa = 3$, $\varepsilon_0 = 8.85 \times 10^{-12}\ C^2\ N^{-1}m^{-2}$

The capacitance is

$$C = \frac{\kappa \varepsilon_0 A}{d}$$

$$= \frac{3 \times 8.85 \times 10^{-12} \times 1}{1 \times 10^{-3}} = 26.55 \times 10^{-9}\ F$$

(ii) If the potential difference between the plates is $V = 120$ V, the magnitude of the charge Q on each plate is,

$$Q = C\,V$$

$$= 26.55 \times 10^{-9} \times 120$$

$$= 3.19 \times 10^{-6}\ C = 3.19\ \mu C$$

(iii) Electric field intensity E between the plates is

$$E = \frac{V}{d}$$

$$= \frac{120}{1 \times 10^{-3}} = 1.2 \times 10^{5}\ V\,m^{-1}$$

Reviewed Exercise

1. When the distance between the two parallel plates of a capacitor is doubled, by what percent does its capacitance change?
2. When the distance between two parallel plates having the charges of equal magnitude and opposite signs is reduced, what will happen to the capacitance and potential difference of the plates?
3. When an insulating material is inserted between the conductors of a capacitor in a vacuum, does its capacitance increase or decrease? Explain.

Key Words: parallel-plate capacitor, dielectric constant, surface charge density, permittivity, relative permittivity

9.4 ENERGY STORED IN A CAPACITOR

A capacitor stores electrical energy in the form of an electric field. We shall now calculate the energy stored by a capacitor.

When the capacitor is charged, the charging process can be viewed as the apparent transfer of charge between two conductors. The charge has been transferred from a conductor at lower potential to a conductor at higher potential. Work has been done for such a transfer of charge. The magnitude of the charge on the two conductors increases gradually and the potential difference between them also increases gradually. The energy stored by the capacitor can be analyzed from the Q-V graph. Figure 9.5 shows the proportionality between the voltage and the charge of a capacitor.

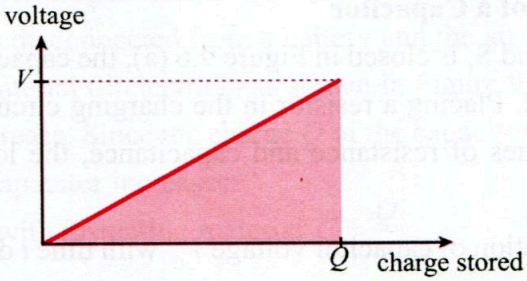


Figure 9.5 The graphical analysis of energy stored in a capacitor

The work done in charging the capacitor is given by the shaded area under the curve (area of the shaded triangle). This work done is equal to $\frac{1}{2}QV$ which is the energy stored by the capacitor. This energy can also be calculated using the average potential difference of the capacitor during charging process.

Before a capacitor is charged each of its conductors has no charge, and the potential difference between two conductors is zero.

Suppose that after the capacitor is charged each conductor receives a charge of magnitude Q and the potential difference between the conductors is V.

If the average potential difference of the capacitor before and after it is charged is $\bar{V}$, then

$$\bar{V} = \frac{0+V}{2} = \frac{V}{2} \qquad (9.10)$$

The work done W for transferring charge of Q between the two conductors is,

$$W = \bar{V}\,Q$$

$$= \frac{1}{2}VQ$$

This amount of work is, in fact, the electrical energy stored by the capacitor in the form of an

electric field.

Therefore the energy of the capacitor is $\frac{1}{2}VQ$.

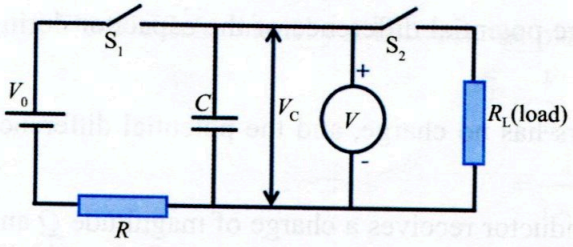
Since $C = Q/V$, the energy, W, of the capacitor can be expressed as

$$\left.\begin{aligned} W &= \frac{1}{2}VQ \\ W &= \frac{1}{2}CV^2 \\ W &= \frac{1}{2}\frac{Q^2}{C} \end{aligned}\right\} \qquad (9.11)$$

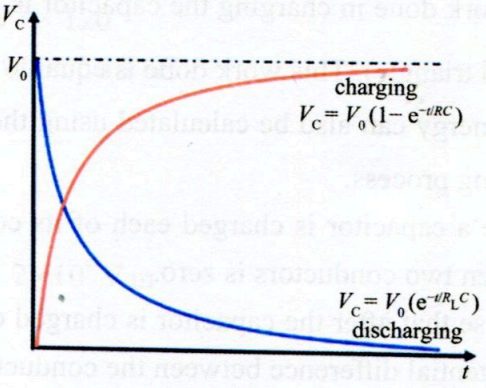
Charging and Discharging of a Capacitor

When the switch S_2 is open and S_1 is closed in Figure 9.6 (a), the capacitor is charged by the power supply through the resistor R. Placing a resistor in the charging circuit slows down the charging process. The greater the values of resistance and capacitance, the longer it takes to charge the capacitor.

Figure 9.6 (b) shows the variation of capacitor voltage V_C with time t during charging process. The voltage rises with time according to the equation $V_C = V_0(1-e^{-t/RC})$. V_0 is the voltage provided by the power supply. The product of the resistance R and the capacitance C is called the time constant τ of a capacitor charging or discharging process. It is measured in second.



(a)



(b)

Figure 9.6 (a) The circuit diagram for charging and discharging a capacitor
(b) Voltage characteristics of charging and discharging a capacitor

After being charged, as described above, the capacitor can then be discharged through the load resistor R_L by opening the switch S_1 and closing S_2 of Figure 9.6 (a). The variation of capacitor voltage with time during this discharging process is shown in Figure 9.6(b). The voltage of the capacitor falls off with time according to the equation $V_C = V_0(e^{-t/R_LC})$.

Example 9.3 What potential difference must be applied across a 10 μF capacitor if it is to have an energy content of 1 J?

Capacitance $C = 10\ \mu F = 10 \times 10^{-6}$ F, energy of capacitor $W = 1$ J,

$$W = \frac{1}{2} C V^2$$

$$V^2 = \frac{2W}{C}$$

$$= \frac{2 \times 1}{10 \times 10^{-6}} = 20 \times 10^4$$

$$V = 447.2 \text{ V}$$

Example 9.4 A charged capacitor of capacitance $C = 35\ \mu F$ is discharged through a resistor of resistance $R = 120\ \Omega$. (i) What is the time constant τ of the discharging process? (ii) What is the elapsed time when the voltage falls to 10 % of its original value?

$C = 35\ \mu F = 35 \times 10^{-6}$ F, $R = 120\ \Omega$

(i) The time constant $\tau = R\,C = 120 \times 35 \times 10^{-6} = 4.2 \times 10^{-3}$ s $= 4.2$ ms

(ii) After a time t the voltage across the capacitor will be $V_C = V_0 (e^{-t/R_L C})$. Since the voltage falls to 10 % of its original value, $V_C = 0.1\ V_0$

$$0.1\ V_0 = V_0 (e^{-t/\tau})$$

$$e^{-t/\tau} = 0.1$$

$$\ln e^{-t/\tau} = \ln 0.1$$

$$= 2.303 \log_{10} 0.1$$

$$-\frac{t}{\tau} = -2.303$$

$$t = 2.303 \times 4.2 \times 10^{-3}$$

$$= 9.673 \times 10^{-3} \text{ s}$$

$$= 9.67 \text{ ms}$$

Reviewed Exercise

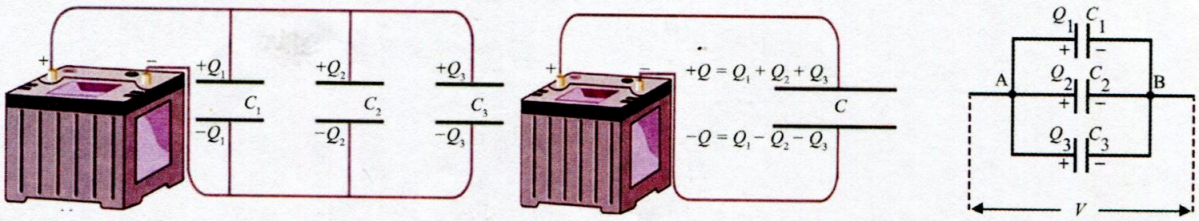
1. A parallel-plate capacitor of capacitance C is given the charge Q and then disconnected from the circuit. How much work is required to pull apart the plates of this capacitor to twice their original separation?
2. In capacitor charging and discharging, why is a resistor required to connect with the capacitor?

Key Words: electrical energy, work, charging and discharging, time constant

9.5 COMBINATION OF CAPACITORS

Several capacitors can be connected together to be used in a variety of applications. There are two types of combination namely; (i) parallel and (ii) series.

(i) Capacitors in Parallel



(a) (b)

Figure 9.7 Capacitors in parallel (a) schematic diagrams (b) in symbols

In Figure 9.7 three capacitors are connected in parallel. The plates of the capacitors connected together at the point A have positive charges and the plates connected together at the point B have negative charges. All the capacitors have the same potential difference, but they carry different amounts of charges.

The respective capacitors have capacitances C_1, C_2 and C_3 and charges Q_1, Q_2 and Q_3 respectively. If the potential difference of each capacitor is V,

$$C_1 = \frac{Q_1}{V} \ , \ C_2 = \frac{Q_2}{V} \ , \ C_3 = \frac{Q_3}{V}$$

If Q is the total charge on the three capacitors, then

$$Q = Q_1 + Q_2 + Q_3$$

$$Q = V(C_1 + C_2 + C_3)$$

$$\frac{Q}{V} = C_1 + C_2 + C_3 \qquad \text{(i)}$$

We can replace the three capacitors in parallel with a single capacitor provided that it has the same total charge Q when the same potential difference V is applied. Such a capacitor is called an equivalent capacitor and its capacitance is the equivalent capacitance of the three capacitors in parallel.

If the capacitance of the equivalent capacitor is C, then

$$C = \frac{Q}{V} \qquad \text{(ii)}$$

From (i) and (ii),

$$C = C_1 + C_2 + C_3 \qquad (9.12)$$

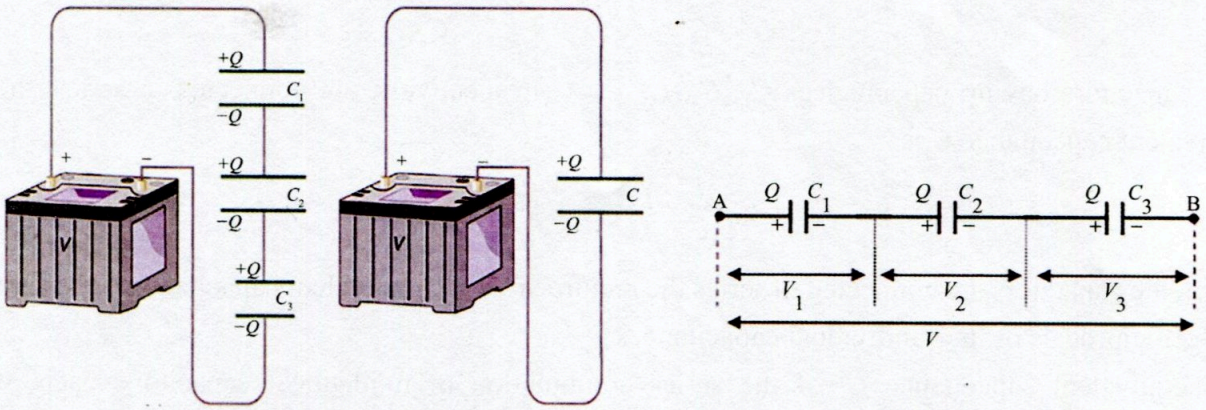
If n capacitors, having capacitances C_1, C_2, C_3, …, C_n and charges Q_1, Q_2, Q_3,…, Q_n respectively, are connected in parallel the equivalent capacitance C is

$$C = C_1 + C_2 + C_3 + \ldots + C_n \qquad (9.13)$$

The equivalent capacitance of the capacitors connected in parallel is the sum of the capacitances of the individual capacitors.

The equivalent capacitance C_p of the parallel combination of n identical capacitors, each of capacitance C, is $C_p = n\,C$.

(ii) Capacitors in Series



(a) (b)

Figure 9.8 Capacitors in series (a) schematic diagrams (b) in symbols

In Figure 9.8 three capacitors having the capacitances C_1, C_2 and C_3 respectively, are connected in series. The negatively charged plate of one capacitor is connected to the positively charged plate of the other.

Since the capacitors are connected in series each capacitor has the same charge. But they have different potential differences.

If the charge on the individual capacitors is Q and their potential differences are V_1, V_2 and V_3, then their capacitances are:

$$C_1 = \frac{Q}{V_1}\ ,\ C_2 = \frac{Q}{V_2}\ ,\ C_3 = \frac{Q}{V_3}\ .$$

Suppose that V is the total potential difference of the capacitor combination.

$$V = V_1 + V_2 + V_3$$

$$= \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3}$$

$$= Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}\right) \qquad \text{(i)}$$

As in the case of capacitors in parallel, we can replace the three capacitors in series with an equivalent capacitor. The charge of the equivalent capacitor is Q when the total potential difference V is applied.

If the capacitance of the equivalent capacitor is C, then

$$C = \frac{Q}{V} \qquad \text{(ii)}$$

From (i) and (ii)

$$\frac{Q}{C} = Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}\right)$$

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \qquad (9.14)$$

If n capacitors having capacitances C_1, C_2, C_3, ... C_n respectively, are connected in series, the equivalent capacitance C is

$$\frac{1}{C} = \left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + ... + \frac{1}{C_n}\right) \qquad (9.15)$$

When the capacitors are connected in series the reciprocal of the equivalent capacitance is the sum of the reciprocals of their individual capacitances.

The equivalent capacitance C_s of the series combination of n identical capacitors, each of capacitance C, is $C_s = \frac{C}{n}$.

Example 9.5 If two capacitors having the capacitances of 4 μF and 12 μF are connected in series, find the equivalent capacitance of the combination of the two capacitors. If the potential difference of the combination is 200 V, find the potential difference of 12 μF capacitor.

$C_1 = 4$ μF, $C_2 = 12$ μF, $V = 200$ V

The equivalent capacitance C of the combination of the two capacitors

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

$$= \frac{1}{4} + \frac{1}{12}$$

$$C = 3 \text{ μF}$$

Suppose that magnitude of the charge on the combination is Q.

$$Q = CV$$

$$= 3 \times 10^{-6} \times 200$$

$$= 600 \times 10^{-6} = 600 \text{ μC}$$

This magnitude of charge is received by the individual capacitors. Therefore, if the potential difference of 12 μF capacitor is V_2.

$$V_2 = \frac{Q}{C_2}$$

$$= \frac{600}{12} = 50 \text{ V}$$

Reviewed Exercise

1. In which connection of the capacitors has each capacitor the same charge?
2. In which connection of the capacitors is the potential difference of each capacitor the same?

Key Words: equivalent capacitance, potential difference, total charge

SUMMARY

A **capacitor** is an electrical device that stores electrical energy in the form of an electric field.

A **parallel-plate capacitor** is the simplest capacitor. It consists of two parallel metal plates separated by air or other insulating material.

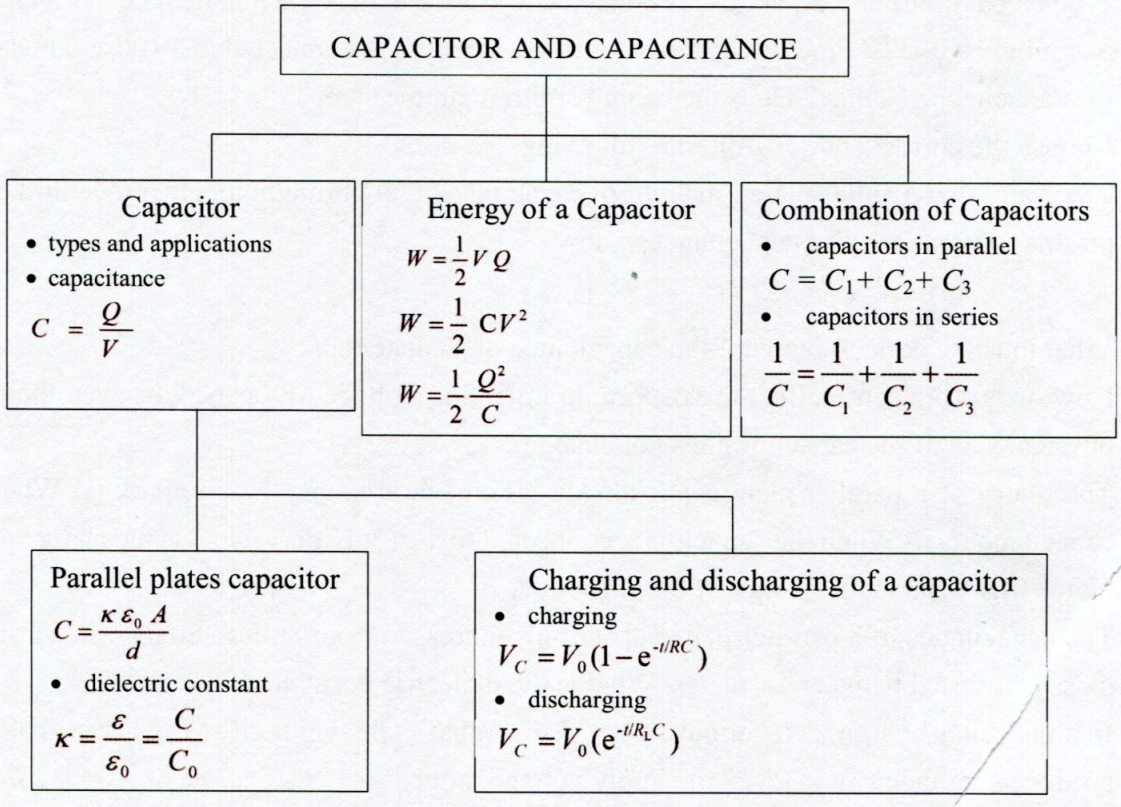
The ratio of the capacitance of a capacitor with an insulating material between its two conductors to the capacitance of that capacitor with a vacuum between its two conductors is called the **dielectric constant** of that insulating material.

EXERCISES

1. State whether the following are True (T) or False (F),
   (i) If one conductor of a capacitor has a charge $+Q$ the other has a charge $-Q$.
   (ii) The charge of the capacitor is the magnitude of the charge on each conductor.
   (iii) The potential difference of the capacitor is twice the potential difference between two conductors of the capacitor.
2. When a parallel-plate capacitor is connected to a 50 V battery each plate receives a charge of magnitude 0.002 C. Find its capacitance. If the capacitor is connected to a 100 V battery, does the capacitance change? Does the charge still remain the same?
3. Choose the correct answer from the following.
   The plates of a parallel-plate capacitor of capacitance C are brought together to a third of their original separation. The capacitance is now
   A. $\frac{1}{9}C$ B. $\frac{1}{3}C$ C. $3C$ D. $9C$
4. What must be done to increase the capacitance of a capacitor?
5. How does a dielectric affect the maximum operating voltage of a capacitor even though the physical size of the capacitor does not change?
6. The plates of a parallel-plate capacitor arc 50 cm^2 in area and 1 mm apart, (i) What is its capacitance? (ii) When the capacitor is connected to a 24 V battery, what is the charge on each plate? (iii) What is the energy of the capacitor?
7. The capacitance of a parallel-plate capacitor is increased from 8 μF to 50 μF when a sheet of glass is inserted between its plates. What is the dielectric constant of the glass?
8. In a capacitor charging RC circuit, $C = 50$ μF. What is the value of resistance R that would produce a voltage rise to 20 % of supply voltage after 1 s?

9. Three capacitors have capacitances of 5 μF, 10 μF and 15 μF.
   (i) Find the equivalent capacitance when they are connected in parallel.
   (ii) Find the equivalent capacitance when they are connected in series.
10. Find the capacitance that can be obtained by combining three 10 μF capacitors in all possible ways.
11. The equivalent capacitance is 10 μF when n identical capacitors are connected in parallel and 0.4 μF when they are connected in series. Determine n and the capacitance of each capacitor.
12. A 35 μF capacitor is needed, but only 10 μF capacitors are available. How should a minimum number of 10 μF capacitors be connected so that the combination has a capacitance of 35 μF?
13. Three capacitors have capacitances of 3 μF, 10 μF, and 15 μF. How should they be connected to obtain the equivalent capacitances of (i) 2 μF (ii) 9 μF (iii) 12.5 μF?
14. Three capacitors of capacitances 3 μF, 10 μF and 15 μF are connected in series with 100 V battery. What is the charge and the potential difference on each capacitor?
15. A capacitor having a capacitance of 2 μF and a charge of 2000 μC is connected in series with another capacitor having a capacitance of 8 μF and a charge of 1600 μC. (i) Find the potential difference of the individual capacitors prior to the connection. (ii) Find the potential difference of the individual capacitors after the connection.

CONCEPT MAP



CHAPTER 10

ELECTRICAL ENERGY, POWER AND HEATING EFFECT OF ELECTRIC CURRENT

Electrical energy can be transformed into a wide variety of other useful forms of energy. The transformation of electrical energy into heat energy is very useful and important. Many home appliances use the heat generated from such transformation. Electrical energy is provided by source of electromotive force such as batteries and generators. In this chapter we shall discuss the electromotive force, electrical energy, electrical power and some applications of heating effect of electric current.

Learning Outcomes

It is expected that students will

- study the electromotive force and electric circuits.
- understand series and parallel connection of batteries.
- distinguish between electrical energy and power.
- state and apply Joule's law of electricity and heat.
- identify some applications of the heating effect of electric current.

10.1 ELECTROMOTIVE FORCE AND ELECTRIC CIRCUITS

When a potential difference is set up between the two ends of a conductor, a current flows through it. A steady current will flow through the conductor if a steady potential difference is maintained between its ends by using batteries and generators. Batteries and generators are called sources of electromotive force.

The sources of electromotive force convert energy from some other forms into electrical form. For example, a battery converts chemical energy into electrical energy to provide electrical power. In an electric generator, mechanical energy is converted into electrical energy.

Electromotive Force

A source of electromotive force has a positive terminal and a negative terminal as shown in Figure 10.1. The positive charges are sent from the negative terminal to the positive terminal within the source. Alternately, it can be said that the main function of the source is to send negative charges from the positive terminal to the negative terminal within the source. In doing so work has to be done by the source.

Electromotive force, abbreviated as emf, is the amount of energy provided by the battery per unit charge to move round the complete circuit.

It is measured in unit of volt (V), equivalent to joule per coulomb of electric charge (J C^{-1})..Despite its name, electromotive force is not actually a force.

Internal Resistance of a Source and Circuit Equation

Generally, a source of electromotive force has an internal resistance. The internal resistance of a battery depends on its size, chemical properties, life time, and temperature. A battery which has an internal resistance r must be viewed as shown in Figure 10.1(a). This means that a resistor of resistance r must be regarded as being connected in series to the battery. The symbol shown in Figure 10.1(b) also represents a battery which has an internal resistance r.

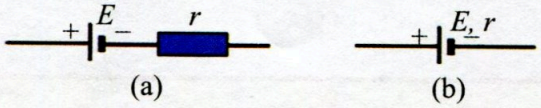


Figure 10.1 Symbols for a battery

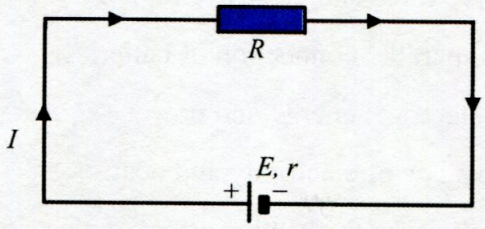


Figure 10.2 A battery connected to an external circuit

In Figure 10.2 a battery of emf E and internal resistance r is connected to a resistor R. A current I flows through the circuit. By Ohm's law, the potential difference across the resistor is $V = I R$. This is the work done in moving a unit positive charge from one end to another of the resistor R. Similarly, the work done in moving a unit positive charge from one end to another of the internal resistance r of the battery is $I r$.

The total work done in moving a unit positive charge round the complete circuit is $I R + I r$, which is the emf E of the battery.

Hence, $$E = I R + I r$$

$$I = \frac{E}{R + r} \qquad (10.1)$$

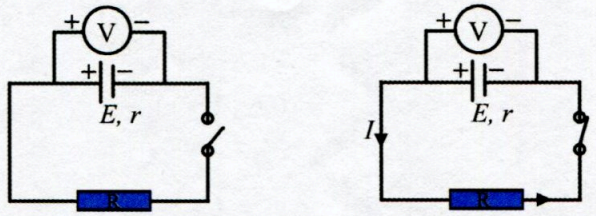
This equation is called the circuit equation.

Eq.(10.1) can be rewritten as, $$I R = E - I r$$

The two ends of the resistor are connected to the positive and negative terminals of the battery as shown in Figure 10.2. Hence, the potential difference between the terminals of a battery when it is connected to an external circuit is given by $V = E - I r$.

When a battery having an internal resistance is used in an electric circuit, the potential difference between the terminals of the battery is always less than its emf. The potential difference across the terminals of a battery connected to an external circuit is called the terminal voltage (or available voltage) of the battery.

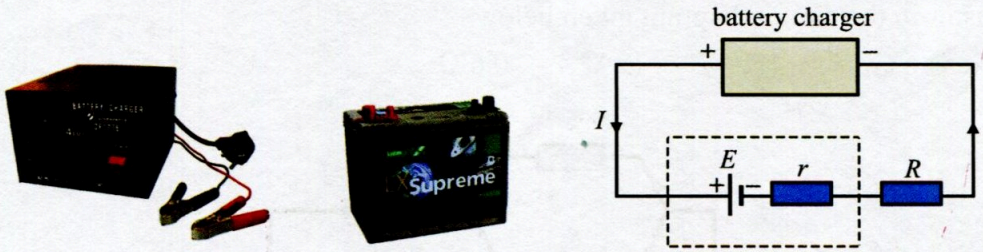
Hence, electromotive force (emf) is the potential difference between the two terminals of a battery or cell in an open circuit (when no current flows) as shown in Figure 10.3 (a). Terminal voltage (available voltage) is the potential difference between the two terminals of a battery or cell in a closed circuit as shown in Figure 10.3 (b). Thus, emf is an open-circuit voltage and terminal voltage is a closed-circuit voltage.



(a) (b)

Figure 10.3 Illustration of (a) emf (b) terminal voltage of a battery

Charging a battery means the chemical energy of the battery which has been used up is now supplied back by electrical energy from some external source (battery charger) shown in Figure 10.4. The external electrical energy required for unit positive charge is equal to the emf E of the battery plus the energy per unit positive charge dissipated in the battery as heat, which is $I\,r$. Therefore, in charging a battery, the potential difference between the terminals is equal to $E + I\,r$.



(a) (b) (c)

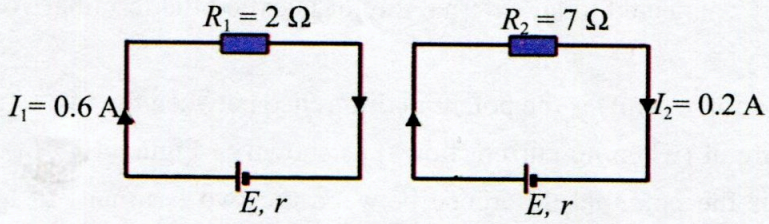
Figure 10.4 (a) Battery charger (b) battery (c) charging circuit

The terminal voltage of a battery when (i) discharging (in use) $V = E - I\,r$, $V < E$ (ii) open circuit (not in use) $V = E$ (iii) charging $V = E + I\,r$, $V > E$.

Example 10.1 When a battery is connected to a 2 Ω resistor it drives a current of 0.6 A through the resistor. When it is connected to a 7 Ω resistor it drives a current of 0.2 A through the resistor. Find the emf and the internal resistance of the battery.

$R_1 = 2\ \Omega, I_1 = 0.6\ \text{A}, R_2 = 7\ \Omega, I_2 = 0.2\ \text{A}$

Let the emf of the battery be E and internal resistance of the battery be r,



(a) (b)

The current in the circuit (a),

$$I_1 = \frac{E}{R_1 + r}$$

$$E = I_1 (R_1 + r)$$

$$= 0.6\ (2 + r) \qquad (1)$$

The current in the circuit (b),

$$I_2 = \frac{E}{R_2 + r}$$

$$E = I_2 (R_2 + r)$$

$$= 0.2\ (7 + r) \qquad (2)$$

From Eq. (1) and (2),

$$0.6\ (2 + r) = 0.2\ (7 + r)$$

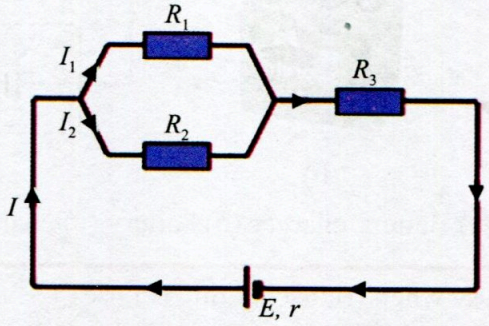
$$0.4\ r = 0.2$$

$$r = 0.5\ \Omega$$

Substituting the value of r into Eq. (1), $E = 0.6\ (2 + 0.5) = 1.5\ \text{V}$

Example 10.2 Find the current flowing through each resistor and the potential difference across the 1 Ω resistor in the circuit diagram given below.

$R_1 = 4\ \Omega,\ R_2 = 6\ \Omega,\ R_3 = 1\Omega,\ E = 12\ \text{V},\ r = 0.6\ \Omega$



If the equivalent resistance of R_1 and R_2 is R_p,

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

$$R_p = 2.4\ \Omega$$

If the equivalent resistance of R_p and R_3 is R,

$$R = R_p + R_3 = 2.4 + 1 = 3.4\ \Omega$$

If the current in the circuit is I,

$$I = \frac{E}{R + r}$$

$$= \frac{12}{3.4 + 0.6}$$

$$= \frac{12}{4} = 3\ \text{A}$$

The potential difference across R_3 is V_3,

$$V_3 = IR_3 = 3 \times 1 = 3\ \text{V}$$

The potential difference between the ends of the combination of R_1 and R_2 resistors is V_p,

$$V_p = IR_p = 3 \times 2.4 = 7.2\ \text{V}$$

Since R_1 and R_2 resistors are connected in parallel, the same potential difference of 7.2 V appears across each resistor.

The current flowing through R_1,

$$I_1 = \frac{V_P}{R_1} = \frac{7.2}{4} = 1.8\ \text{A}$$

The current flowing through R_2,

$$I_2 = \frac{V_P}{R_2} = \frac{7.2}{6} = 1.2\ \text{A}$$

Example 10.3 What is the difference between electromotive force (emf) and potential difference (voltage)?

Electromotive force (emf) is the amount of energy, provided by the battery, per unit charge to move round the complete circuit.

The potential difference or voltage is the amount of energy required to move a unit charge from one end to another end of a circuit element.

Reviewed Exercise

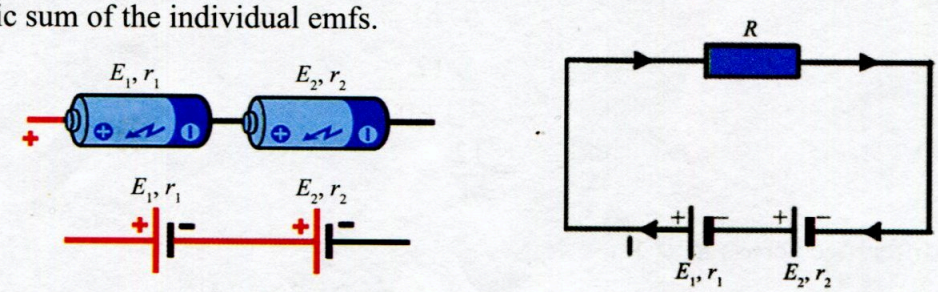
1. What is meant by terminal voltage (available voltage) of a battery and how is it related to emf?
2. What are (i) open circuit and (ii) closed circuit? Illustrate with diagrams.

Key Words: electromotive force (emf), available voltage, internal resistance, open circuit, closed circuit, current

10.2 BATTERIES IN SERIES AND IN PARALLEL

1. Batteries in Series

When two or more electromotive sources (batteries) are connected in series, the resultant emf is the algebraic sum of the individual emfs.

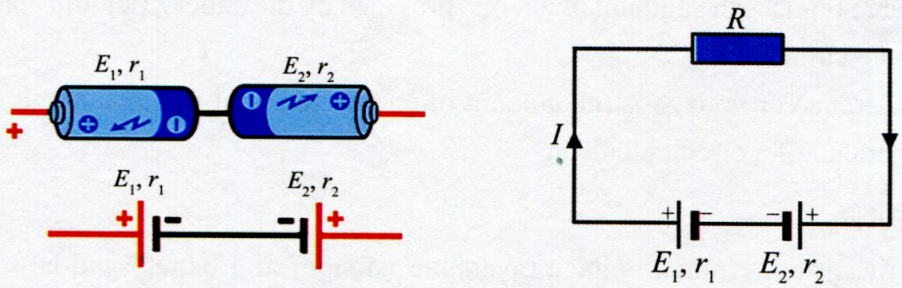


(a) (b)

Figure 10.5 (a) Batteries in series aiding
(b) batteries in series aiding connected with an external resistor

In Figure 10.5 (a) two batteries are connected in series aiding. In this arrangement the currents leaving the batteries are in the same direction so that the resultant emf is $E_1 + E_2$. Such a connection of batteries is called series aiding. The total internal resistance of those two batteries is $(r_1 + r_2)$. They can be regarded as a single battery having an emf $(E_1 + E_2)$ and internal resistance $(r_1 + r_2)$. In Figure 10.5 (b) two batteries in series aiding arrangement are connected to an external resistor (resistance R). The current in the circuit is

$$I = \frac{E_1 + E_2}{R + r_1 + r_2} \qquad (10.2)$$



(a) (b)

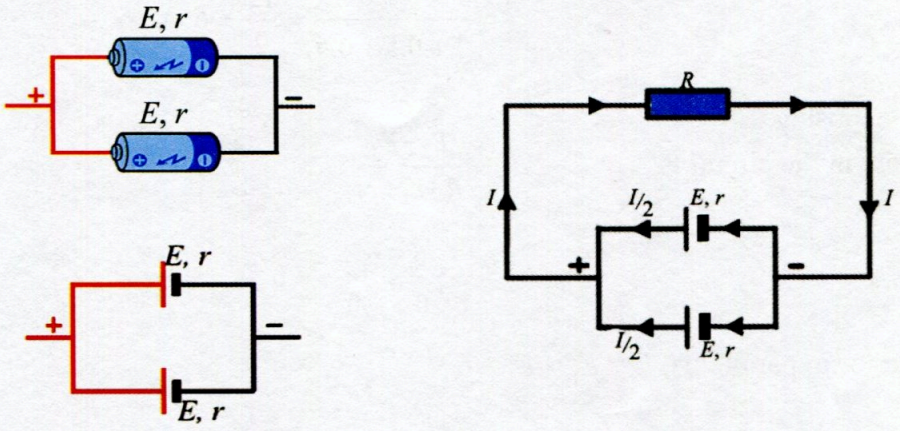
Figure 10.6 (a) Batteries in series opposing
(b) batteries in series opposing connected with an external resistor

In Figure 10.6 (a) two batteries are connected in series opposing. The emfs of the batteries are in opposition. Thus, the resultant emf is the difference between the individual emfs. Such a connection of batteries is called series opposing. If E_1 is greater than E_2 the resultant emf of those two batteries is $(E_1 - E_2)$. But their total internal resistance is still $(r_1 + r_2)$. Thus, they can be regarded as a single battery having an emf $(E_1 - E_2)$ and an internal resistance $(r_1 + r_2)$.

In Figure 10.6 (b) two batteries in series opposing arrangement are connected to an external resistor (resistance R). The current in the circuit is

$$I = \frac{E_1 - E_2}{R + r_1 + r_2} \qquad (10.3)$$

2. Batteries in Parallel



(a) (b)

Figure 10.7 (a) Batteries in parallel
(b) batteries in parallel connected with an external resistor

In Figure 10.7 (a) two batteries having equal emfs and equal internal resistances are connected in parallel. It should be noticed that the same polarity terminals of the batteries are connected together. The resultant emf of the parallel-combination is just the emf E of a single battery. Since the internal resistances of the batteries are in parallel, the resultant internal resistance is $\frac{r}{2}$. This combination can be regarded as a single battery having an emf E and an internal resistance $\frac{r}{2}$.
In Figure 10.7 (b), the two batteries in parallel combination are connected to an external resistor (resistance R). In the circuit the resultant emf is E and the total resistance is $(R + \frac{r}{2})$.

$$I = \frac{E}{R + \frac{r}{2}} \qquad (10.4)$$

Since the two batteries have equal emfs and equal internal resistances the current flowing through each battery is $\frac{I}{2}$.

In parallel connection of batteries, care must be taken that the terminal of the same polarity should be connected together. Otherwise, batteries may be damaged.

Example 10.4 Two batteries each having an emf of 6 V and an internal resistance of 0.5 Ω are connected (i) in series and (ii) in parallel. Find the current in each case when the batteries are connected to a 1 Ω resistor.

(i) in series aiding $E_1 = E_2 = 6$ V, $r_1 = r_2 = 0.5\ \Omega$, $R = 1\ \Omega$

The current in the circuit is, $$I = \frac{E_1 + E_2}{R + r_1 + r_2}$$

$$= \frac{6+6}{1+0.5+0.5} = \frac{12}{2} = 6 \text{ A}$$

(ii) in series opposing

The current in the circuit is, $$I = \frac{E_1 - E_2}{R + r_1 + r_2}$$

$$= \frac{6-6}{1+0.5+0.5} = 0$$

(iii) The batteries in parallel is, $E = E_1 = E_2 = 6$ V

The current in the circuit is, $$I = \frac{E}{R + \frac{r}{2}}$$

$$= \frac{6}{1 + \frac{0.5}{2}} = 4.8 \text{ A}$$

Reviewed Exercise

1. What are the advantages and disadvantages of (i) the series aiding connection of batteries (ii) the parallel connection of batteries?
2. A battery has an emf of 6 V and an internal resistance of 0.5 Ω. How many batteries are necessary to pass a current of 1 A through a 22 Ω resistor in an electric circuit?

Key Words: internal resistance, series aiding, series opposing, parallel combination, resultant emf

10.3 ELECTRICAL ENERGY AND ELECTRICAL POWER

Electrical Energy

Electrical energy is energy associated with the flow of charge through any part of circuit. Electrical energy can be generated from source of emf such as batteries, generators, dynamos and photovoltaic cells. It can also be stored in other forms for future application using fuel cells, batteries, capacitors and inductors.

(a) (b)

Figure 10.8 Conversion of electrical to heat energy in the resistor

In Figure 10.8, a load of resistance R is connected to a battery, which is a source of emf. As the current I flows through R, the potential at the point A is higher than that at the point B. Suppose that the potential difference between A and B is V. By the definition of potential difference, the work done in bringing a unit positive charge from A to B is V.

Therefore, the work done W in bringing the amount of charge Q from A to B is

$$W = Q\,V$$

Suppose the amount of charge Q passes through a cross-sectional area of the resistor R in the time t.

$$I = \frac{Q}{t} \quad \text{(or)} \quad Q = I\,t$$

The work done W is obtained as, $W = V\,I\,t$

The work done W by the battery in taking the charge Q from A to B is, in fact, the electrical energy supplied by the battery. This electrical energy is transformed into heat energy in the resistor R. It is because the electrons collide with the atoms in the resistor R when they pass through it. Hence the atoms acquire additional energy and therefore heat energy is produced.

Since the potential difference between A and B is $V = I\,R$, the work done (or) the electrical energy W produced by the battery is

$$\left.\begin{aligned} W &= V\,I\,t \\ W &= I^2\,R\,t \\ W &= \frac{V^2}{R}t \end{aligned}\right\} \qquad (10.5)$$

If an electric motor, instead of electric lamp, is connected between A and B in Figure 10.8 (a) the electrical energy will be transformed into mechanical energy.

Unit of Electrical Energy

The practical unit of electrical energy is kilowatt hour (kWh). The relation between the unit of electrical energy kWh and the unit of work J is

$$1\ \text{kWh} = 1000\ \text{W} \times 1\ \text{h}$$
$$= 1000 \times 60 \times 60 = 3.60 \times 10^6\ \text{J}$$

Example 10.5 If a current of 2 A flows through a 50 Ω resistor for 30 min find the amount of electrical energy dissipated in the resistor.

$R = 50\ \Omega,\ I = 2\ \text{A},\ t = 30\ \text{min} = \frac{1}{2}\ \text{h}$

Let W be the electrical energy dissipated in the resistor,

$$W = I^2 R t$$
$$= 2^2 \times 50 \times \frac{1}{2}$$
$$= 100\ \text{Wh} = 0.1\ \text{kWh}$$

Example 10.6 An electric lamp of 60 Ω connected to a 240 V mains line is used for 45 min. (i) Find the amount of electrical energy dissipated in the lamp. (ii) Find the cost of using it if electricity costs 35 kyats per unit.

$R = 60\ \Omega,\ V = 240\ \text{V},\ t = 45\ \text{min} = \frac{3}{4}\ \text{h}$

(i) Let W be the amount of electrical energy dissipated in the lamp.

$$W = \frac{V^2}{R} t$$
$$= \frac{(240)^2}{60} \times \frac{3}{4}$$
$$= 720\ \text{Wh} = 0.72\ \text{kWh}$$

(ii) One unit of electricity = 1 kWh

The cost of using 0.72 kWh = $35 \times 0.72 = 25.2$ kyats

Electrical Power

Electrical power is the rate of transfer of electrical energy. If the electrical energy W is transferred in the time t, then the electrical power P is,

$$P = \frac{W}{t}$$

Since $W = V I t = I^2 R t = \frac{V^2}{R} t$

$$P = V I = I^2 R = \frac{V^2}{R} \qquad (10.6)$$

Unit of electrical power

The unit of electrical power P is the watt (W). If 1 joule (J) of electrical energy is transferred in 1 second (s) the electrical power is 1 joule per second (J s^{-1}) or 1 watt (W).

Example 10.7 If a 1200 W electric iron is used for 50 min, by how many units does the electricity meter reading increase? Calculate the payment if one unit of electricity costs 35 kyats.

$P = 1200$ W, $t = 50$ min $= \frac{5}{6}$ h

Let W be the electrical energy used by the electric iron.

$$P = \frac{W}{t}$$

$$W = P\,t = 1200 \times \frac{5}{6}$$

$$= 1000 \text{ Wh} = 1 \text{ kWh}$$

The electricity meter reading increases by 1 uint of electricity.

One unit of electricity = 1 kWh

Therefore, the payment = $1 \times 35 = 35$ kyats

Reviewed Exercise

1. Which quantity determines the rate at which electrical energy is delivered by a current?
2. When an air conditioner is connected to a 220 V mains line it draws a current of 10 A and is used for 6 h. (i) Find the amount of electrical energy consumed by it. (ii) Calculate the cost of using it if the electrical energy costs 50 kyats per unit.

Key Words: potential difference, electrical energy, electrical power, electricity meter

10.4 JOULE'S LAW OF ELECTRICITY AND HEAT

When electric current passes through a conductor, the electrons in the conductor are accelerated due to electric field. Therefore, they collide with the other atoms and lose their kinetic energy which appears heat energy.

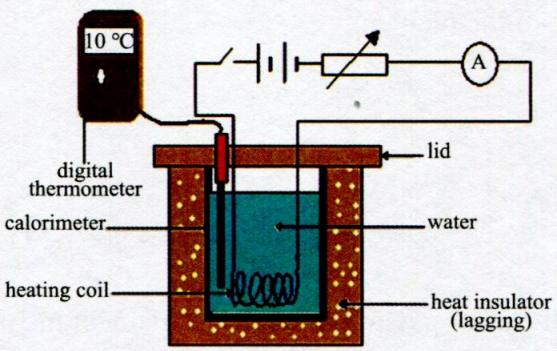


Figure 10.9 Joule's calorimeter

English physicist James Prescott Joule in 1841 experimentally studied the heating effect of electric current. Heat energy is produced in the resistor R (heating coil) as shown in Figure 10.9. The electrical energy W supplied by the battery is transformed into heat energy H in the resistor R. The electrical energy W (measured in joule) is related to the amount of heat H (measured in calorie) as follows.

$$W = J H \qquad (10.7)$$

where J is a constant called Joule's mechanical equivalent of heat ($J = 4.2$ J cal^{-1}).

When $W = I^2 R t$ is substituted in equation (10.7) the following equation is obtained.

$$H = \frac{I^2 R t}{J} \qquad (10.8)$$

This equation represents Joule's law of electricity and heat which can be stated as follows.

The amount of heat produced in a resistor due to a current flowing through it is directly proportional to the square of the current, the value of resistance and the time taken by the current to pass through the resistor.

H can also be written as follows:

$$H = \frac{V I t}{J} = \frac{V^2 t}{R J} \qquad (10.9)$$

Example 10.8 If a 60 W electric lamp is connected to a 220 V mains line find (i) the current in the lamp (ii) the resistance of tungsten wire of the lamp (iii) the amount of charge passing through the filament in 1 min and (iv) the amount of heat produced by the filament in 1 min.

(i) $P = 60$ W, $V = 220$ V

Let I be the current in the filament.

$$P = V I$$

$$I = \frac{P}{V}$$

$$= \frac{60}{220} = 0.27 \text{ A}$$

(ii) Let R be the resistance of the filament.

$$P = \frac{V^2}{R}$$

$$R = \frac{V^2}{p}$$

$$= \frac{220 \times 220}{60} = 806.67\ \Omega$$

(iii) Let Q be the amount of charge passing through the filament in 1 min.

$$I = \frac{Q}{t}$$

$$Q = I t$$

$$= 0.27 \times 60 = 16.2 \text{ C}$$

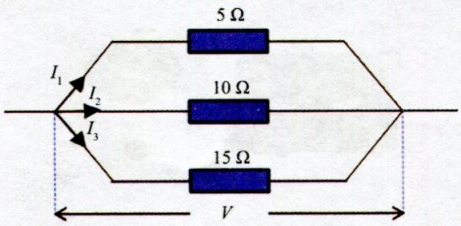
(iv) Let H be the amount of heat generated in 1 min.

$$H = \frac{V I t}{J}$$

$$= \frac{P\ t}{J}$$

$$= \frac{60 \times 60}{4.2} = 857.14 \text{ cal}$$

Example 10.9 One 5 Ω, one 10 Ω and one 15 Ω resistors are connected in parallel. If each resistor has an electrical power rating of 0.5 W, find the maximum potential difference which may be supplied to the parallel combination and the current in each resistor.



Since, the resistors are connected in parallel the potential difference across each resistor is the same. The electrical power dissipated in the 5 Ω resistor (minimum resistance) would be maximum according to $P = \frac{V^2}{R}$. Therefore, the 5 Ω resistor must be used to find the maximum potential difference.

$$0.5 = \frac{V^2}{5}$$

$$V = 1.58\ V$$

The current in the 5 Ω resistor, $I_1 = \frac{V}{R_1}$

$$= \frac{1.58}{5} = 0.32 \text{ A}$$

The current in the 10 Ω resistor, $I_2 = \frac{V}{R_2}$

$$= \frac{1.58}{10} = 0.16 \text{ A}$$

The current in the 15 Ω resistor, $I_3 = \frac{V}{R_3}$

$$= \frac{1.58}{15} = 0.11 \text{ A}$$

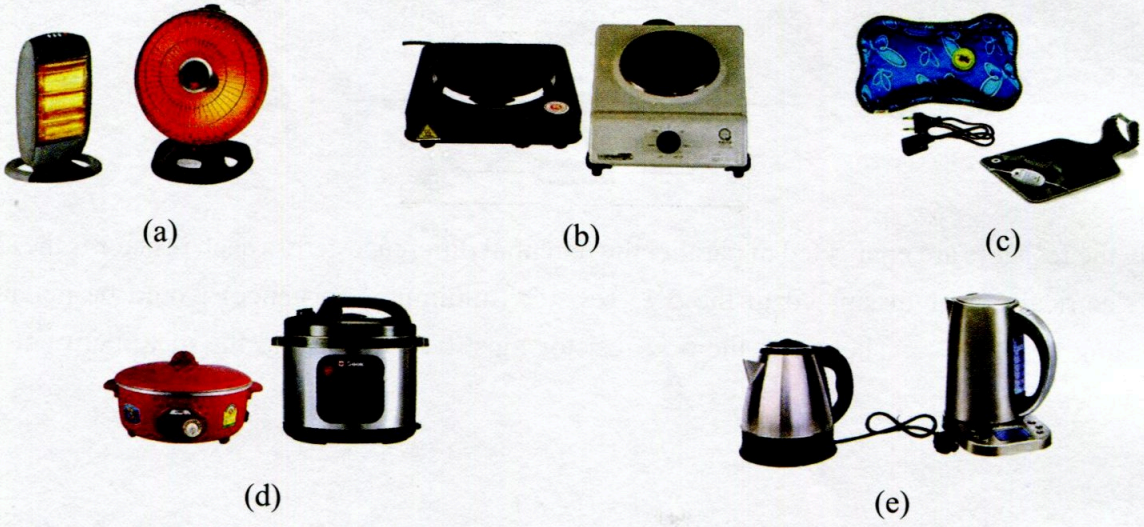
Reviewed Exercise

1. On which factors does the heat produced in a current carrying conductor depend?
2. An electric stove of 1200 W is connected to a 220 V mains line. (i) Find its resistance. (ii) Find the current flowing through it. (iii) Find the amount of calories produced in one second by it. (iv) Find the electrical power produced by it when the voltage of the mains line drops to 180 V.

Key Words: electrical energy, heat energy, Joule's mechanical equivalent of heat, power rating

10.5 SOME APPLICATIONS OF THE HEATING EFFECT OF CURRENT

Electrical energy can be transformed into a wide variety of other useful forms of energy. In general, most of the electrical energy supplied to a resistor in an electric circuit is converted into heat. The heating effect of electric current has special applications in domestic appliances. For example, electric stoves, electric cookers, electric irons, electric kettle, electric heaters and electric heating bag as shown in Figure 10.10. They change electrical energy into heat energy.



(a) (b) (c)

(d) (e)

Figure 10.10 Domestic electrical appliances (a) electric heater (b) electric stove (c) electric heating bag (d) electric cooker (e) electric kettle

Some electrical appliances which use the heating effect of current are described below.

Light bulb

An electric light bulb is made of a glass bulb consisting a tungsten filament which has high melting point (~3000 °C) as shown in Figure 10.11. A small amount of argon, an inert gas, is added after the air is removed. The tungsten filament becomes hot when a current flows through it. At a temperature of 1800 °C it emits white light. Since the inert gas is added the light bulb can be used safely up to the temperature 2500 °C. The inert gas also reduces evaporation of the filament. The electric light bulb converts most of the electrical energy into heat and light energy.

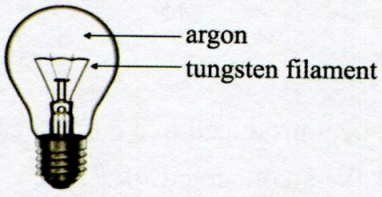
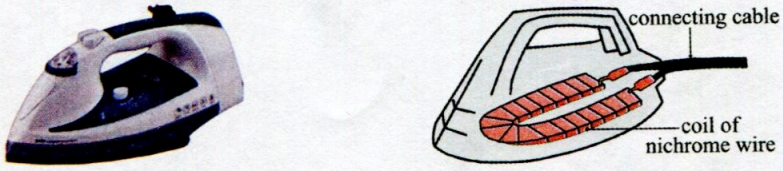


Figure 10.11 Tungsten filament lamp

Electric light bulbs are various sizes and electrical power. A small light bulb used in a torch has a few watts and the bulbs used for illuminating airfields have electrical power of the order of 5000 W. The voltage and the power of a simple light bulb are usually specified as 220 V, 60 W etc.

Electric Iron

Figure 10.12 shows an electric iron. A mica sheet is wound around the nichrome wire and is inserted between two other mica sheets. Mica is a good electrical insulator and can withstand very high temperature. The combination is then placed on a heavy metal plate and is sandwiched with another metal plate. When a current flows through the nichrome wire it produces heat energy. The heating element is controlled by a thermostat that switches the current on and off to maintain the selected temperature. The position of the nichrome wire in the electric iron is shown in Figure 10.12 (b). An electric iron has an electric power of 750-1000 W.



(a) (b)

Figure 10.12 Electric iron

Electric stove

In an electric stove, a coil of nichrome wire is placed in the groove of the heat-resistant material as shown in Figure 10.13. When a current flows through the nichrome coil, it produces heat energy. An electric stove has an electrical power ranging from 600 W to 2000 W.

Figure 10.13 Electric stove

The electrical power of some electrical appliance used in homes are given in Table 10.1.

Table 10.1 The electrical power of some electrical appliance used in homes

Electrical Appliance	Electrical Power (W)
Laptop computer	15 -45
Reading lamp	40 -100
Refrigerator	100 -250
Radio and television receiver	~ 150
Electric iron	750 -1000
Electric stove	~ 1200

Fuse

A fuse is a thin short wire of tin-lead alloy which is used in electric circuits to prevent electrical appliances from being damage if excessive current flows through them. The fuse wire is fitted in a glass tube (cartridge) or on a porcelain block as shown in Figure 10.14. For example, a 3 A fuse is used in the electric circuit. When a current greater than 3 A flows in the circuit, the fuse becomes so hot that it will melt and break the circuit. Thus, the current stops flowing and electrical appliance in the circuit is not damaged. For an appliance rated as 700 W or less, a 3 A fuse should be fitted. For appliances above 700 W, a 13 A fuse should be fitted.

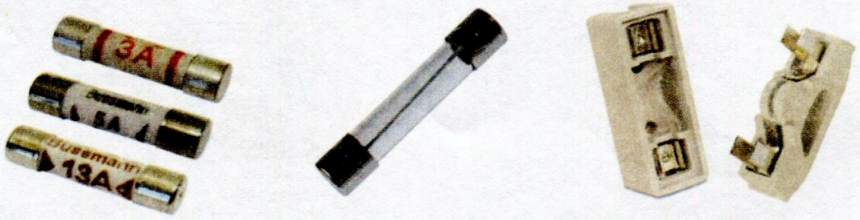


Figure 10.14 Fuses

Example 10.10 A 3 A fuse is used in a circuit which contains a source of 220 V. Find the maximum power which can safely be consumed.

$V = 220$ V, $I = 3$ A

Let P be the maximum power which can safely be consumed.

$$P = VI$$
$$= 220 \times 3 = 660 \text{ W}$$

Reviewed Exercise

1. An electric circuit installed in a house contains a 5 A fuse and the voltage is 220 V. Can twenty 60 W electric lamps be used at the same time in that current?
2. What should the rating of the fuse be used for an electric stove of 1200 W, 220 V?

Key Words: heating effect, electrical appliances, tungsten filament, nichrome wire, mica, fuse

SUMMARY

Electromotive force, abbreviated as emf is the amount of energy, provided by the battery, per unit charge to move round the complete circuit.

The potential difference across the terminals of a battery connected to an external circuit is called the **terminal voltage** (or **available voltage**) of the battery.

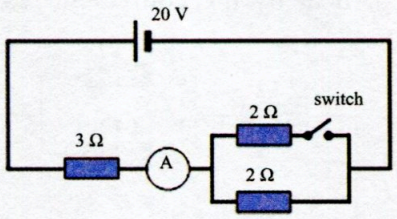
Electrical energy is energy associated with the flow of charge through any part of circuit. **Electrical power** is the rate of transfer of electrical energy.

Joule's law stated as the amount of heat produced in a resistor due to a current flowing through it is directly proportional to the square of the current, the value of resistance and the time taken by the current to pass through the resistor.

A fuse is a thin short wire of tin-lead alloy which is used in electric circuits to prevent electrical appliances from being damage if excessive current flows through them.

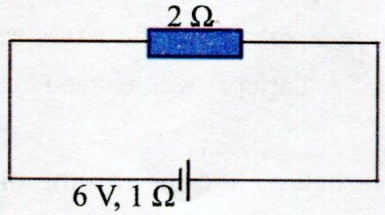
EXERCISES

1. How is electricity produced by an electric cell?
2. In an electric circuit, which component is used for (i) supply electricity, (ii) close (or) open the electric circuit, (iii) providing a path for the flow of electric current?
3. Can the terminal voltage of a battery ever exceed its emf? If so, explain the necessary condition.
4. In the electric circuit shown below, find the reading of the ammeter A when the switch is (i) open (ii) closed. (Neglect the internal resistance of the battery.)

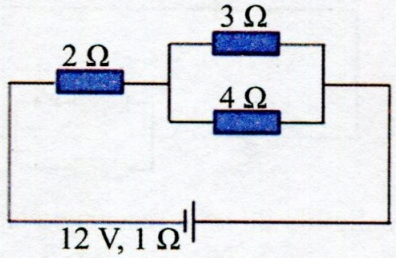


5. When a 12 V battery of negligible internal resistance is connected to a resistor, a current of 3 A flows through it. When another battery of emf 6 V is in the circuit in series with the first one, the current flowing through the resistor remains at 3 A. Find the internal resistance of the second battery.
6. When two 6 V batteries, having the same internal resistance and connected in series, are connected to a 5 Ω resistor, the current in the circuit is 2 A. When these batteries are in parallel, a current of 1.5 A flows through when connected to another resistor. Find the resistance of the resistor.
7. With two identical batteries, how would you arrange the bulbs and batteries in a circuit to get the maximum possible total power out? (Assume the batteries have negligible internal resistance.)
8. What is electrical energy? Express its unit.
9. (i) Define electrical power. (ii) Write down the unit of electrical power. (iii) How many joules are there in 1 kWh?
10. Why do we use fuses in an electric circuit?
11. Two heaters of the same power and the same resistance are used to heat the water. In which of the following cases will the water heat up faster? (i) if heaters are connected in series (ii) if heaters are connected in parallel.
12. Which draws more current, a 100 W light bulb or a 75 W light bulb when connected to a main line? Which has higher resistance?

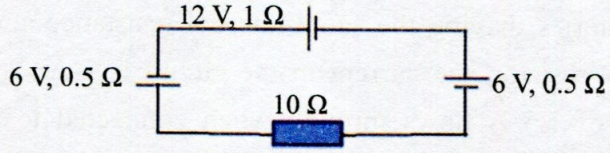
13. State the Joule's law of electricity and heat.
14. Why is electrical energy transformed into heat energy when a current flows through a resistor?
15. An electric iron draws a current of 3 A when it is connected to a 220 V mains line. How many kcal of heat are produced per min?
16. Find the amount of calories produced per second by a 2 Ω resistor in the circuit diagram shown below.



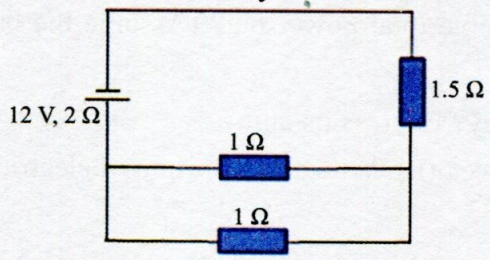
17. Find the rate of production of heat by the equivalent resistance of resistors in the circuit diagram shown below.



18. Find the amount of heat produced in 10 min by a 10 Ω resistor in the circuit diagram shown below.



19. Find the rate of production of heat in the battery in the circuit diagram shown below.



20. Which of the following metals cannot be used as fuse wire?
    A. iron B. silver C. tin-lead alloy
21. Why should the material used for fuse have low melting point?
22. An electric circuit installed in a house contains a 5 A fuse and the voltage is 220 V. Find the maximum electrical power which can safely be used.

23. An electric circuit installed in an office contains a 10 A fuse and the voltage is 220 V. Ten 100 W electric lamps and two 150 W refrigerators are being used there. (i) Find the cost of using all the lamps and two refrigerators for 10 h. (ii) Find the maximum number of 60 W electric lamps which can be safely used in addition. (Assume that electricity costs 50 kyats per unit.)

CONCEPT MAP

ELECTRICAL ENERGY, POWER AND HEATING EFFECT OF ELECTRIC CURRENT

Electromotive Force (emf)

$$I = \frac{E}{R + r}$$

Batteries in series and in parallel

- batteries in series aiding

$$I = \frac{E_1 + E_2}{R + r_1 + r_2}$$

- batteries in series opposing ($E_1 > E_2$)

$$I = \frac{E_1 - E_2}{R + r_1 + r_2}$$

- batteries in parallel

$$I = \frac{E}{R + \frac{r}{2}}$$

Electrical Energy and Electrical Power

- electrical energy

$$W = VIt = I^2Rt = \frac{V^2}{R}t$$

- electrical power

$$P = VI = I^2R = \frac{V^2}{R}$$

Joule's law of electricity and heat

$$W = JH$$

$$H = \frac{I^2Rt}{J}$$

Some applications of the heating effect of current

- light bulb
- electric iron
- electric stove
- fuse

CHAPTER 11

ELECTROMAGNETIC INDUCTION, GENERATION AND DISTRIBUTION OF ELECTRICITY

A changing magnetic field could produce a current. The discovery was made in 1831 by Michael Faraday and Joseph Henry working independently at the same time. The effect is called electromagnetic induction and it is the physics behind the generators that provide the electricity used in our modern society.

Learning Outcomes

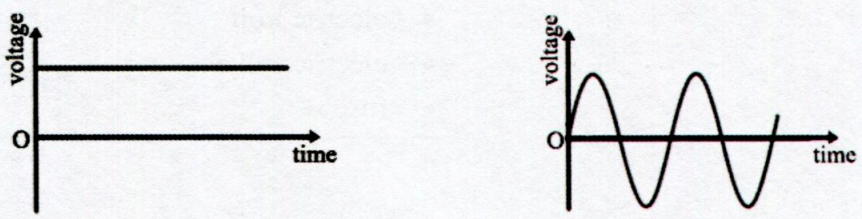
It is expected that students will

- identify two types of electric current.
- investigate electromagnetic induction and its applications.
- understand the generation and distribution of electricity.
- describe the principles of house wiring and explain the dangers associated with electricity.

11.1 ALTERNATING CURRENT

Electric current can be classified as two types: direct current (DC) and alternating current (AC). The direct current is a steady unidirectional current. In a direct current the electrons flow steadily in the same direction all the time. The current that flows in a flashlight, cell phones, laptops, and any appliance running on the batteries is direct current.

Another type of current is of great industrial important. This is the alternating current. An alternating current is a current that reverses its direction at regular time intervals. In the alternating current the direction of the flow of electrons changes periodically. Current flowing in power lines and normal household electricity from a wall outlet (plug) is alternating current. In Japan and in the most of America, the voltage is between 100 V and 120 V, whereas Europe and most other countries in the world (including Myanmar) use a voltage between 220 V and 240 V. The frequency of AC current is 50 Hz/60 Hz in Japan, 60 Hz in US and 50 Hz in Europe and most other parts of the world.



(a) (b)

Figure 11.1 Graphical representations of the magnitude of (a) DC (b) AC voltages as a function of time

Figure 11.1 shows the graphical representations of the magnitude of DC and AC voltages as a function of time.

The method of production of an alternating current is described below. When a coil of wire having quite a few turns is rotated a few cycles per second continuously in a magnetic field, an emf is induced in the coil as shown in Figure 11.2. When a galvanometer is connected across the ends of the rotating coil, the galvanometer will show a peculiar result. In each revolution the current produced in the coil due to the induced emf, rises to a maximum value, then falls to zero, reverses, rises to a maximum and falls to zero again.

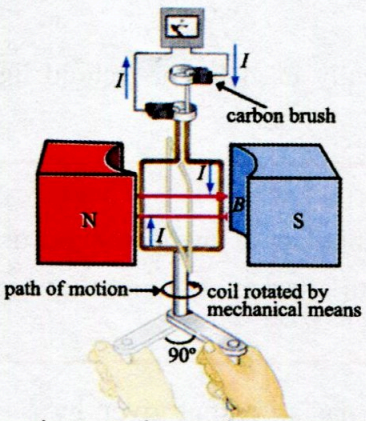


Figure 11.2 Generating an alternating current in a rotating coil

As the coil continues to rotate, this sequence of changes occurs in a periodic manner. The current having this characteristics is, in fact, the alternating current, called induced current. The induced emf also shows such periodic variation. Variation of induced emf is shown in Figure 11.3 for one complete rotation of the coil.

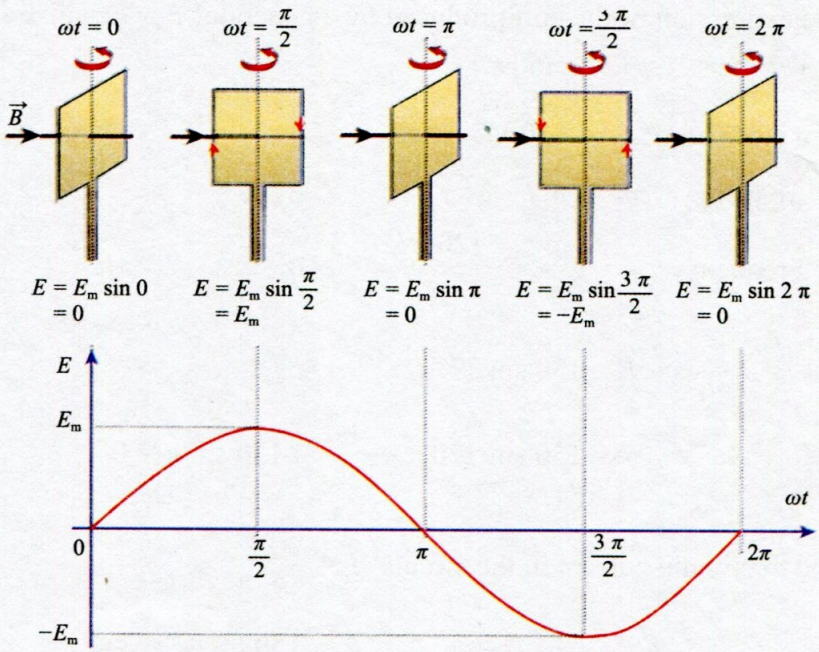


Figure 11.3 Variation of induced emf as a function of ωt

The waveform of the emf shown in Figure 11.3 is called a sine wave. It is represented by

$$E = E_m \sin \omega t \qquad (11.1)$$

where E is the induced emf at a particular instant, E_m is the maximum value of emf and ω is the angular velocity of the rotating coil.

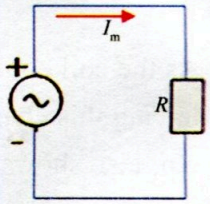
The angular velocity ω is equal to $2\pi f$, where f is the frequency of the emf, (i.e. the number of complete cycles generated in 1 s). Therefore, a rotating coil in a magnetic field is the basic of an AC generator (a device for generating alternating current). The magnitude of an alternating current at any instant varies precisely the same way as shown by Eq. (11.1) and is given by

$$I = I_m \sin \omega t \qquad (11.2)$$

The alternating current is universally used for household lighting, heating, cooking and other domestic functions.

The advantage of alternating current is that it is relatively easy to change the voltage of the current. Furthermore, the inevitable loss of energy that occurs when current is distributed over long distances is far smaller with alternating current (at high voltage and low current) than with direct current.

Example 11.1 An AC generator has an output given by $E = 150 \sin 377\ t$. (i) Find the maximum emf output, the frequency of the source, and the emf at $t = 1/120$ s. (ii) Find the maximum current in the circuit when generator is connected to a 50 Ω resistor in the figure.



$E = 150 \sin 377\ t$

By comparing the expression of the emf produced by a generator $E = E_m \sin \omega\ t$.

$E_m = 150$ V, $\omega = 377$ rad s^{-1} $= 120\pi$ rad s^{-1}

(i) The maximum emf output $E_m = 150$ V

Angular frequency $\omega = 2\pi f$

$$\text{Frequency } f = \frac{\omega}{2\pi} = \frac{120\pi}{2\pi} = 60 \text{ Hz}$$

$$\text{The emf at } t = \frac{1}{120} \text{ s, } E = 150 \sin 377\ t$$

$$= 150 \sin 120\pi \times \frac{1}{120} = 150 \sin \pi = 0$$

(ii) $R = 50\ \Omega$, The maximum current in the circuit $I_m = \dfrac{E_m}{R}$

$$= \frac{150}{50} = 3 \text{ A}$$

Example 11.2 The current in an AC circuit at any time t seconds is given by $I = 20 \sin (100 \pi t)$ A. Find (i) the peak value (maximum value) of current, the period and the frequency (ii) the value of the current when t = 0 (iii) the value of the current when t = 8 ms and (iv) the time when the current reaches 10 A.

By comparing the expression of $I = I_m \sin \omega t$.

$$I = 20 \sin (100 \pi t)$$

(i) Peak value $I_m = 20$ A, $\omega = 100 \pi$

$$\text{Period} \quad T = \frac{2\pi}{\omega} = \frac{2\pi}{100\pi} = 0.02 \text{ s} = 20 \text{ ms}$$

$$\text{Frequency} \quad f = \frac{1}{T} = \frac{1}{0.02} = 50 \text{ Hz}$$

(ii) When $t = 0$, $I = 20 \sin (0) = 0$

(iii) When $t = 8$ ms, $I = 20 \sin [100 \pi (0.008)]$

$= 11.76$ A

(iv) When $I = 10$ A, $10 = 20 \sin (100 \pi t)$

$0.5 = \sin (100 \pi t)$

$t = 0.0017 \text{ s} = 1.7 \text{ ms}$

Reviewed Exercise

1. What is the major difference between AC and DC current?
2. Why AC current is better than DC in long distance distribution of electrical power?
3. Is the current from the outlet of telephone charger, AC or DC?

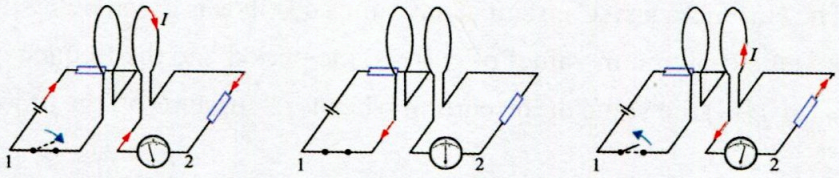
Key Words: direct current, alternating current, flow of electrons, induced emf, waveform, angular velocity, frequency, angular frequency

11.2 ELECTROMAGNETIC INDUCTION

Faraday's Experiments of Induced Current

In 1831, Joseph Henry and Michael Faraday independently discovered the methods of producing electric current from magnetic field. Faraday's experiments are described below.

The first wire loop is connected to a battery through a switch and the second wire loop is connected to a galvanometer as shown in Figure 11.4. As soon as the switch is closed, a current flows through the first loop and the galvanometer gives a brief deflection as shown in Figure 11.4 (a). That is, a momentary current flows through the second loop. When the switch is kept closed as shown in Figure 11.4 (b), the galvanometer shows no deflection (there is no induced current in the second loop). However, when the switch is opened again, the galvanometer shows the brief deflection in the other direction as shown in Figure 11.4 (c), which indicates that the induced current flows in the opposite direction.



(a) (b) (c)

Figure 11.4 Induced current produced by a wire loop carrying current

Again, a momentary current flows through the second loop (Q) when the first loop (P) carrying a current is moved abruptly towards or away from the second loop as shown in Figure 11.5.

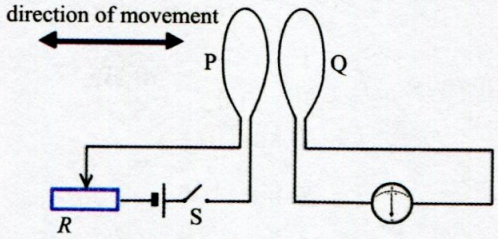


Figure 11.5 Induced current produced by a moving coil carrying current

Moreover, the same thing happens when a magnet, instead of the current-carrying first loop, is moved abruptly towards or away from the second loop as shown in Figure 11.6.

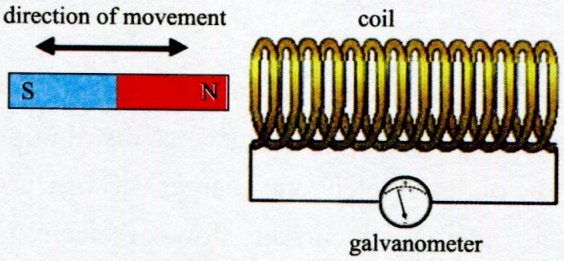


Figure 11.6 Induced current produced by a moving magnet

In the above experiments, switching the current in the first loop ON and OFF (or) the motion of the current-carrying loop (or) the motion of the magnet; all causes a change in the magnetic lines of force through the second loop and causes a momentary current to flow through it.

That is, a current flows through a loop of wire whenever there is a change in the magnetic lines of force through the loop. This phenomenon is called electromagnetic induction. As mentioned above, this current is called an induced current and the electromotive force (emf) which causes the induced current to flow is called an induced emf.

Magnetic Flux Density and Magnetic Flux

The magnetic flux density or magnetic induction B is the number of magnetic lines of force passing perpendicularly through a unit area of material. Magnetic flux density B is related to magnetic field strength (magnetic field intensity) H by

$$B = \mu H \qquad (11.3)$$

where μ is the permeability of the medium in which the magnetic field exists. The permeability of free space (vacuum) $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1} = 1.257×10^{-6} H m^{-1}.

The SI unit of magnetic flux density is tesla (T) which is equivalent to weber per metre squared (Wb m^{-2}).

Magnetic flux ϕ_B is a measurement of the total magnetic lines of force which passes through a surface area. In electromagnetism, the magnetic flux through a surface is the surface integral of the normal component of the magnetic field B over that surface.

$$\phi_B = \int \vec{B} \cdot d\vec{A} \qquad (11.4)$$

where $d\vec{A}$ is the elemental surface area.

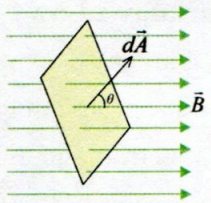
If the magnetic field is uniform over a plane area A, the magnetic flux ϕ_B passing through that area A is simply

$$\phi_B = \int \vec{B} \cdot d\vec{A} = \int B \cos\theta \, dA = B \cos\theta \int dA$$

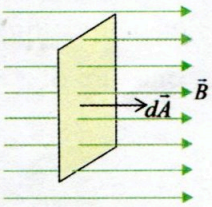
$$\phi_B = B\,A\cos\theta \qquad (11.5)$$

where θ is the angle between the magnetic field $\vec{B}$ and $d\vec{A}$.

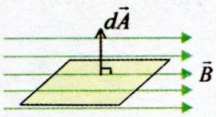
When the field is subtended an angle θ to the surface area, $\phi_B = B\,A\cos\theta$.



When the field is perpendicular to the surface area, $\theta = 0°$.
Hence, $\phi_B = B\,A\cos 0° = B\,A$



When the field is parallel to the surface area, $\theta = 90°$.
Hence, $\phi_B = B\,A\cos 90° = 0$ as shown in figure.



The SI unit of magnetic flux is the weber (Wb). The weber is named after the German physicist Wilhelm Eduard Weber (1804-1891). 1 Wb = 1 V s = 1 kg m^2 s^{-2} A^{-1}. Table 11.1 shows the SI units of magnetic parameters.

Table 11.1 The SI unit of B, H, ϕ

Quantity	SI unit
Magnetic flux density B	tesla (T)
Magnetic field strength H	ampere per metre (A m^{-1})
Magnetic flux ϕ	weber (Wb)

Magnetic field strength and magnetic flux density are vector quantities whereas magnetic flux is a scalar quantity.

Example 11.3 A square loop of wire 10 cm on a side is in a magnetic field of 1.25 T. What are the maximum and minimum values of flux that can pass through the loop?

Area of square loop $A = 10 \times 10 = 100\ \text{cm}^2 = 0.01\ \text{m}^2$

The magnetic flux passing through the area A is $\phi_B = B\,A \cos\theta$

When the plane of loop is perpendicular to $\vec{B}$, it is a maximum for $\theta = 0°$.

$$\phi_B = B\,A \cos 0° = 1.25 \times 0.01 \times 1 = 0.0125\ \text{Wb}$$

When the plane of loop is aligned with $\vec{B}$, the minimum value occurs for $\theta = 90°$.

$$\phi_B = B\,A \cos 90° = 0$$

Reviewed Exercise

- Differentiate between the magnetic flux and magnetic flux density.

Key Words: electromagnetic induction, induced current, induced emf, magnetic flux, magnetic flux density

11.3 FARADAY'S LAW AND LENZ'S LAW

Faraday's Law

In Faraday's experiments mentioned above, the change in magnetic flux through a wire loop produces an induced emf. Based on these experiments, Michael Faraday (1831) discovered the law of electromagnetic induction stated as below.

The magnitude of induced emf in a wire loop is directly proportional to the rate of change of magnetic flux through that loop.

For one turn of a wire loop, induced emf is

$$E = -\frac{d\,\phi_B}{d\,t}$$

For a wire loop with N identical turns,

$$E = -N\frac{d\,\phi_B}{d\,t} \qquad (11.6)$$

In the above equation, the minus sign gives the direction of the induced emf.

If the wire loop is not moving (or) is moving parallel to the magnetic field lines, there is no induced emf (or) induced current.

Lenz's Law

Faraday's law gives the magnitude of the induced emf and Lenz's law gives its direction. In 1834, the Russian physicist Heinrich Friedrich Emil Lenz stated a law as follows.

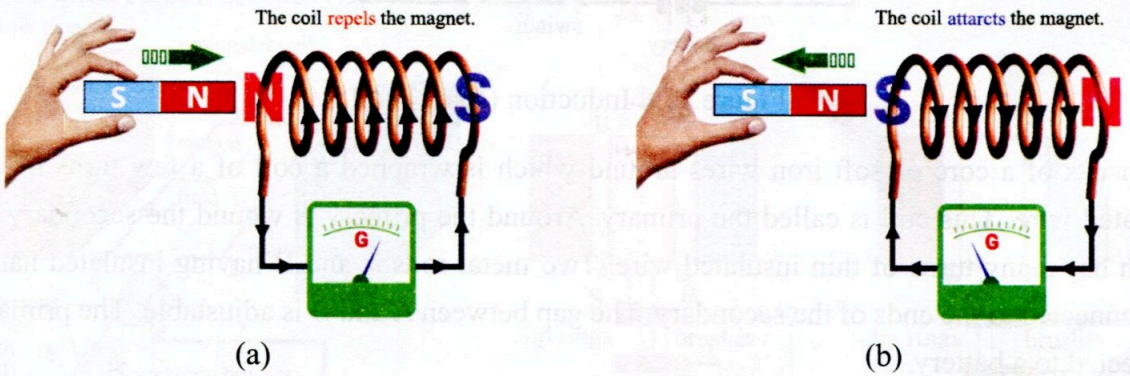
An induced emf is always in a direction that opposes the change in the original magnetic flux that causes it.

The direction of induced current can be found by applying Lenz's law. The change in magnetic flux is, in fact, a decrease or an increase in the magnetic lines of force.

Lenz's law is a good example of the principle of conservation of energy. It shows how the energy from the mechanical work done against the opposing force experienced by the moving magnet or rotating coil is transformed into electrical energy.

In Figure 11.7 (a), when a magnet is moved towards the coil, the magnetic flux through the coil increases. The induced current in the coil must produce a magnetic field which opposes an increase in the original magnetic flux. This means that the direction of the magnetic field produced by the induced current must be towards the magnet. In order to be so the end of the coil nearer the magnet must act as a north pole. Therefore, the current in the coil must be flowing in a counter clockwise direction if viewed from the magnet according to right-hand rule.

If the magnet is moved away from the coil the induced current will flow in the opposite direction as shown in Figure 11.7(b). The direction of the induced current can be found as explained above.



(a) (b)

Figure 11.7 (a) When N pole of the magnet is moved towards the coil, the induced current flows in the counter clockwise direction.

(b) When N pole of the magnet is moved away from the coil, the induced current flows in the clockwise direction if viewed from the magnet.

Reviewed Exercise

1. Why is Faraday's law important?
   (Hint: how changing magnetic fields can cause the flow of current in wires.)
2. What does negative sign indicate in Faraday's law of electromagnetic induction formula?

Key Words: magnetic flux, magnetic lines of force, induced current, induced emf

11.4 APPLICATIONS OF ELECTROMAGNETIC INDUCTION

1. Induction Coil

One of the useful applications of electromagnetic induction is an induction coil. In 1851, a German instrument maker named Heinrich Ruhmkorff showed how an electric spark 30 cm long could be produced by using a battery of only a few volts. The apparatus he used is called an induction coil or Ruhmkorff spark coil. The construction of an induction coil is shown in Figure 11.8.

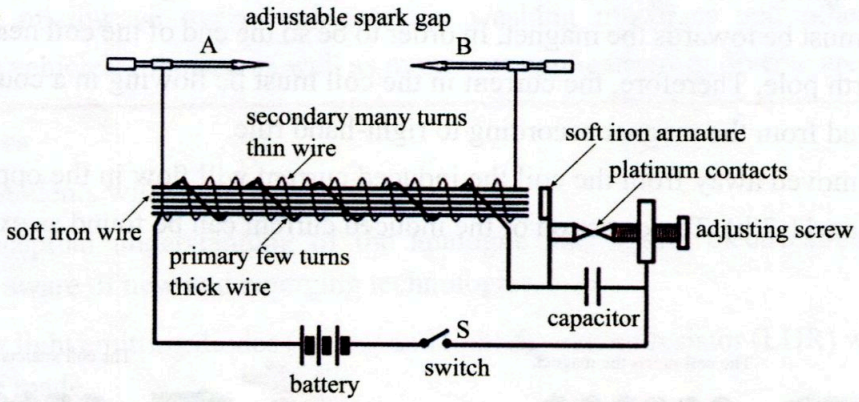


Figure 11.8 Induction (spark) coil

It consists of a core of soft iron wires around which is wrapped a coil of a few turns of thick insulated wire. This coil is called the primary. Around the primary is wound the secondary coil, which has many turns of thin insulated wire. Two metal rods A and B having insulated handles are connected to the ends of the secondary. The gap between A and B is adjustable. The primary is connected to a battery.

When the switch S is pressed a current flows through the primary. As soon as the current flows in the primary an induced emf develops in every turn of the secondary. Since the secondary consists of many turns, the induced emf is very high indeed.

As the current flows through the primary, the soft iron core becomes an electromagnet which attracts a soft iron armature mounted on the spring. When this happens, the platinum contacts are separated and hence the current stops flowing. As soon as the current stops flowing, very high induced emf develops in the secondary again.

Since there is no current in the primary, the soft iron core becomes demagnetized. At that moment the armature is returned by the spring to its original position. The contact is remade again and the whole process is repeated. When the primary is connected to a 6 V battery a voltage of about 30 000 V can develop across the terminals A and B. When the terminal of A and B are kept a small distance apart, sparks are formed between them. A capacitor is connected across the platinum contacts to prevent sparking and wearing away of contacts.

The induction coil is used in the ignition system of motor cars. The primary of an induction coil is connected to a 6 V or 12 V battery and the secondary is connected to a sparking plug. As a very high voltage is set up between the two terminals of the plug, sparking occurs and the mixture of gases is ignited.

2. AC Generator

A generator is an electromagnetic device which converts mechanical energy into electrical energy. Most generators give out alternating current. AC generators are also called alternators.

AC generators work on the principle of Faraday's law of electromagnetic induction. A simple AC generator is shown in Figure 11.9 (a). It consists of a rectangular coil mounted on an axle which is fixed between the poles of a permanent magnet. When the coil is rotated, the magnetic field through it changes. This induces a current in the coil. Using Fleming's right-hand rule, the induced current flows from A to B, and C to D in the coil.

This induced current flows through the two slip rings (S_1 and S_2) and then through the external load from P to Q as shown in Figure 11.9 (b). The two slip rings (S_1 and S_2) each make sliding contact with two fixed carbon brushes (B_1 and B_2), respectively.

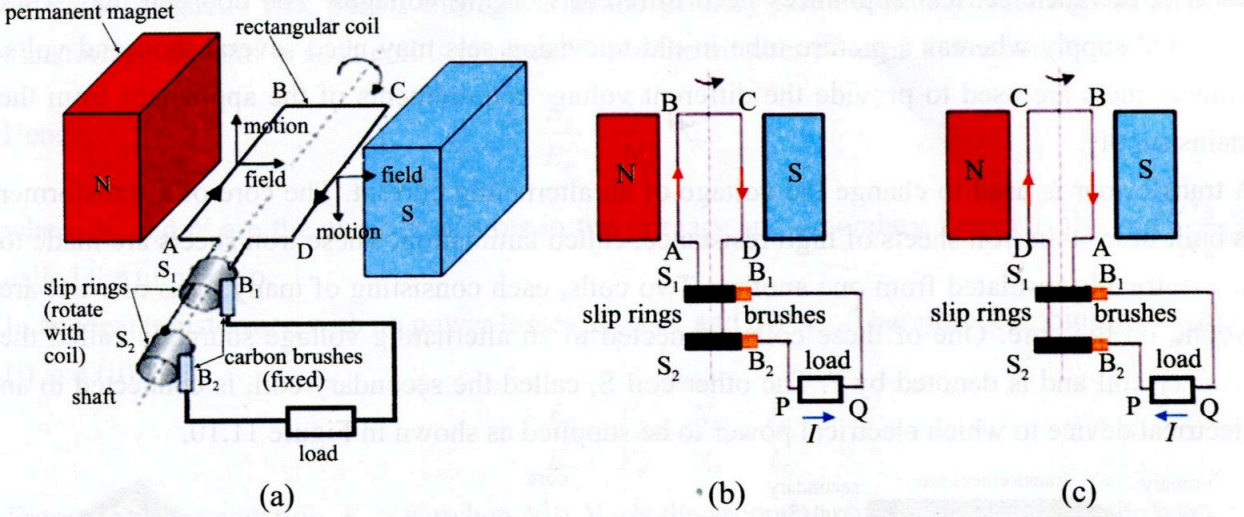


Figure 11.9 The working principle of an AC generator

When the coil has rotated through 180° the sides of the coil will have changed places. Now the current flows from B to A and D to C in the coil, and from Q to P in the external load as shown in Figure 11.9 (c). Every time the coil turns through 180° the current reverses its direction. This is an alternating current which changes its direction when the coil is vertical.

If the speed of rotating coil is doubled, both the frequency of the alternating current and the rate of cutting of the magnetic lines of force will be doubled. This means that the maximum output voltage is also doubled. When the number of turns in the coil is doubled, the frequency of the output voltage is the same although its magnitude is doubled.

The induced electromotive force of a generator can be increased by (i) increasing the speed of

rotation of the coil, (ii) increasing the area, (iii) the number of turns in the coil, and (iv) using stronger magnets.

The shaft (axle) of the coil is rotated by an external force such as falling water or an engine. If the wire loop is rotating with a constant angular velocity ω, $\theta = \omega t$ and $\phi_B = B A \cos \omega t$.

The induced emf is $E = -\dfrac{d\phi_B}{dt} = -BA\dfrac{d}{dt}(\cos \omega t) = B A \omega \sin \omega t$

For a coil of N loops, $E = N B A \omega \sin \omega t = E_m \sin \omega t$ (11.7)

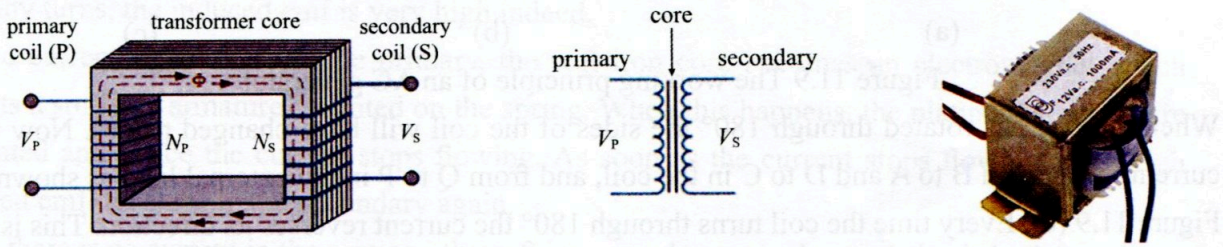
where $E_m = N B A \omega$.

There is no structural difference between a generator and an electric motor. The difference lies in the way energy is converted. In an electric motor, electrical energy is used to rotate the coil to provide mechanical energy. In the case of generators, mechanical energy is used to rotate the coil to produce electrical energy.

3. Transformer

One of the another useful application is the construction of transformers that are commonly used today. Different electrical appliances need different working voltages. The doorbell may work on a 6 V supply whereas a picture tube in old television sets may need several thousand volts. Transformers are used to provide the different voltage requirements of the appliances from the mains supply.

A transformer is used to change the voltage of an alternating current. The core of a transformer is built up of thin iron sheets of high resistance, called lamination. These iron sheets are made to be electrically insulated from one another. Two coils, each consisting of many turns of wire, are wound on the core. One of these coils connected to an alternating voltage source is called the primary coil and is denoted by P. The other coil S, called the secondary coil, is connected to an electrical device to which electrical power to be supplied as shown in Figure 11.10.



(a) (b) (c)

Figure 11.10 (a) Simple transformer (b) its symbol (c) actual photograph of a transformer

When an alternating voltage is applied to the primary P, an alternating current flows through it. The alternating current in the primary produces an alternating magnetic flux in the core and nearly all the flux passes through the secondary S. This changing magnetic flux produces an induced emf and

an induced current in the secondary. Hence, an alternating voltage is obtained from the secondary. The voltage in the secondary and the voltage in the primary have the same frequency.

The use of laminated iron (mumetal) core is to minimize the power loss due to the induced current flowing in the core. There is also a power loss due to heating of wire used in the coils when the current flows through them. In a systematically constructed transformer, the power losses are so small that the power input can be assumed to be equal to the power output.

Suppose that V_p is the voltage applied to the primary, and V_s is the voltage obtained from the secondary. If I_p is the current flowing in the primary, and I_s is that flowing in the secondary, then

power input = power output

$$V_p I_p = V_s I_s \qquad \text{(i)}$$

As the product of voltage and current is the same on both coils of a transformer, it follows that a transformer which increases the voltage will reduce the current in the same proportion, and vice versa.

Suppose that E_p and E_s are the induced emfs in the primary and secondary, respectively. Since the same magnetic flux passes through both the primary and secondary windings in an ideal transformer, the emf in each coil is proportional to the number of turns in the coil.

Hence,
$$\frac{E_S}{E_P} = \frac{N_S}{N_P} \qquad \text{(ii)}$$

where N_p and N_s are the number of turns in the primary and secondary, respectively, and $\frac{N_S}{N_P}$ is called the turns ratio.

In an ideal transformer, with no power losses, $E_p = V_p$ and $E_s = V_s$. Therefore, combining equations (i) and (ii)

$$\frac{E_S}{E_P} = \frac{V_S}{V_P} = \frac{N_S}{N_P} = \frac{I_P}{I_S} \qquad (11.8)$$

From the above equation, $V_s > V_p$ when $N_s > N_p$. If the secondary has more turns than the primary, the voltage obtained from the secondary is greater than the voltage applied to the primary. Such a transformer is called a step-up transformer.

On the other hand, $V_s < V_p$ when $N_s < N_p$. If the secondary has fewer turns than the primary, the voltage obtained from the secondary is smaller than the voltage applied to the primary. Such a transformer is called a step-down transformer.

In a transformer, the changing magnetic field not only induces currents in the secondary coil, but also currents in the iron core itself. These currents flow in little circles in the iron core and are called eddy currents.

The efficiency of a transformer is the ratio of the output power to the input power as usual. Most of the transformers have full load efficiency from 95% to 98.5%. Because of power losses, there is no ideal transformer with 100% efficiency.

To improve the efficiency of a transformer, the following features should be taken into consideration.

- Soft magnetic material should be used to make the core.
- A laminated core should be used to reduce the flow of eddy current.
- A special core design should be used to ensure that the magnetic field from the primary coil completely links with the secondary coil.
- Low-resistance copper wires should be used as coils to reduce energy loss in the form of heat.

Transformers work on AC, but not on DC. Unless there is a changing current in the input coil, no voltage is induced in the output coil.
Note that, a transformer connected to a DC supply can be damaged as the high current flowed through the input coil can make it overheat.

Example 11.4 The generator coil with 200 turns is rotated through one-fourth of a revolution in 15 ms. The turn of circular coil has a 5 cm radius and is in a uniform magnetic field of 1.25 T. Find the induced emf in 15 ms.

$N = 200$, $t = 15$ ms $= 15 \times 10^{-3}$ s, $r = 5$ cm $= 0.05$ m, $B = 1.25$ T

Area of loop $A = \pi r^2 = 3.142 \times (0.05)^2 = 7.86 \times 10^{-3}\,m^2$

Since the area of the loop and the magnetic field are constant, change in magnetic flux due to rotational angle from 0° to 90° (one-fourth of a revolution).

$$\phi = B A \cos\theta$$

The change in magnetic flux is

$$\Delta\phi = B A \Delta(\cos\theta)$$
$$= 1.25 \times 7.86 \times 10^{-3} \times (\cos 90° - \cos 0°) = -\ 0.0098 \text{ Wb}$$

The induced emf $$E = -N\frac{\Delta\phi_B}{\Delta t}$$

$$= -\ 200 \times \frac{(-0.0098)}{15 \times 10^{-3}} = 130.7 \text{ V}$$

Example 11.5 A step-down transformer is used to light a 12 V, 24 W lamp from the 240 V mains. The current through the primary is 125 mA. What is the efficiency of the transformer?

$V_p = 240$ V, $V_s = 12$ V, output power $P_{out} = 24$ W, $I_p = 125$ mA $= 0.125$ A

Input power $P_{in} = V_p I_p$

$$= 240 \times 0.125 = 30 \text{ W}$$

$$\text{Efficiency} = \frac{\text{output power}}{\text{input power}} \times 100\ \%$$

$$= \frac{24}{30} \times 100\ \% = 80\ \%$$

Example 11.6 A 2 kW transformer is used to step down an AC supply of 240 V to 60 V. (i) What is the turns ratio of this transformer? (ii) What will the output current be if there is a 10 % power loss in the transformer? (iii) How large is the output current if there is no power loss (i.e. 100 % efficiency)?

$V_p = 240$ V, $V_s = 60$ V, input power $P_{in} = 2$ kW $= 2000$ W

(i) Turns ratio $$\frac{N_S}{N_P} = \frac{V_S}{V_P} = \frac{60}{240} = \frac{1}{4}$$

(ii) If there is a 10 % power loss,

output power = 90 % of input power

$$60 \times I_S = \frac{90}{100} \times 2000$$

$$I_S = \frac{90}{100} \times \frac{2000}{60} = 30 \text{ A}$$

The output current is 30 A if there is a 10 % power loss.

(iii) If there is no power loss,

output power = input power

$$V_s I_s = V_p I_p$$

$$60 \times I_s = 2000$$

$$I_s = 33.33 \text{ A}$$

The output current is 33.33 A if there is no power loss.

Reviewed Exercise

1. What devices must be used to deliver (i) thousands of voltage at the terminals of a sparking plug of motorcar engine using a 12 V battery (ii) 12 V AC from 220 V AC mains?
2. Explain each of the following: (i) A transformer will not work on DC. (ii) The core of a transformer needs to be laminated. (iii) If a transformer increases voltage, it reduces current.
3. (i) Which type of transformer must be used to change a high current and a low alternating voltage into a lower current and a higher alternating voltage? (ii) Which type of transformer must be used to change a low current and a high alternating voltage into a higher current and a lower alternating voltage? (iii) Which type of transformer is used in a welding machine? Why?

Key Words: induction coil, primary coil, secondary coil, lamination, mumetal, eddy current, step-down transformer, step-up transformer, turn ratio

11.5 POWER TRANSMISSION

The electricity generated at the power station is transmitted to the consumers through the power cables. For instance, a generator produces 10 MW of electrical power at 10 kV. The current supplied can be calculated as follows.

$$P = VI$$

Thus the current, $$I = \frac{P}{V} = \frac{10 \times 10^6}{10 \times 10^3} = 1000 \text{ A}$$

This is a large amount of current. If it is transmitted through the power cables, which may have a resistance of 1 ohm per kilometre, then the amount of electrical power lost as heat in 1 km of transmission line will be given by

$$\text{Power loss} = I^2 R = (1000)^2 \times 1 = 1 \text{ MW}$$

This is equivalent to almost 10 % of power generated at the power station. Power loss is one of the main problems in the long distance transmission of electricity.

One way to reduce power loss is to use very thick cables of low resistance. These cables would, however, be very heavy and uneconomical. A more practical way would be to reduce the magnitude of the transmission current by stepping up the voltage. In the above example, if the electrical power is transmitted at 20 kV, then

Transmission current, $$I = \frac{P}{V} = \frac{10 \times 10^6}{20 \times 10^3} = 500 \text{ A}$$

Hence, power loss per km $= I^2 R = (500)^2 \times 1 = 250\,000 \text{ W} = 250 \text{ kW}$

The power loss would only about 2.5 % of the power generated at the power station. When increased the voltage 10 kV to 20 kV at the power station, the power loss will be reduced to 4 times.

The national grid system is used to distribute electricity around the country. An advantage of the grid system is power stations in areas where the demand is low can be used to supply areas where the demand is high. First, an alternating current is generated at high voltage (e.g. 10 kV) from the power station. The voltage is then stepped up (e.g. 230 kV) for transmission and finally stepped down at the substations in stages (e.g. 66 kV, 22 kV, 11 kV, 6.6 kV, 440 V, 230 V) before it is used in factories and homes. This process can be carried out with the help of transformers. Power distribution through a grid system is illustrated in Figure 11.11.

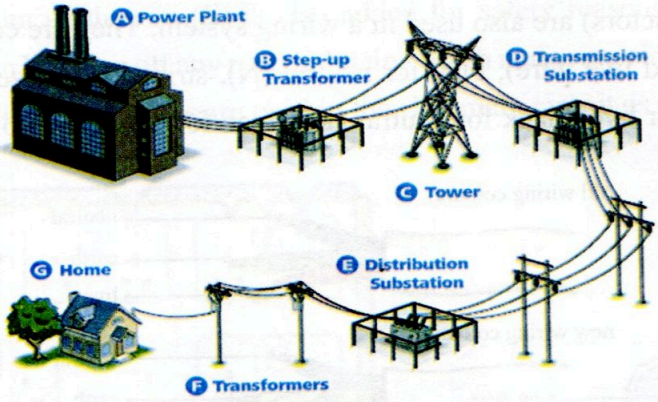


Figure 11.11 Illustration of electrical power distribution

Reviewed Exercise

- Explain why AC rather than DC is used for transmitting electrical power.

Key Words: electrical power, transmission, power loss, power cables, national grid system

11.6 PRINCIPLES OF HOUSE WIRING

Now we will study the proper way of using electricity in a home, that is house wiring system. Electricity supplied to the house is a single-phase form of alternating current (AC), carried by a thick cable which contains two wires which are well insulated from each other. One of these wires is the live wire (L) and the other is the neutral wire (N). The current enters the house through one wire and returns to the local substation through the other. The live wire (L) is a dangerous wire as it carries a high voltage while the neutral wire (N) is usually at zero volt.

The electricity is fed into the house through the main fuse, meter, main switch and then to the distribution box as shown in Figure 11.12. From the distribution box there are normally three circuits; namely the lighting circuit, the power circuit (ring main circuit) and heating circuit (e.g. cooker circuit) for high power cookers.

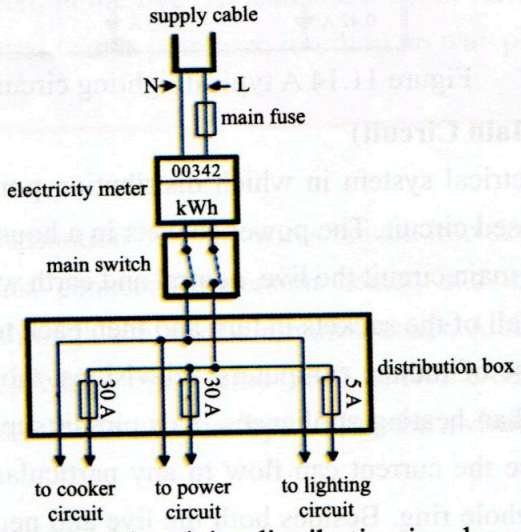


Figure 11.12 Electricity distribution in a house

Three cables (or conductors) are also used in a wiring system. They are colour coded with brown for live (L) (also called live wire), blue for neutral (N), striped green-yellow for earth (E). The old colour code: red for live, black for neutral and green for earth, is also still in use as shown in Figure 11.13.

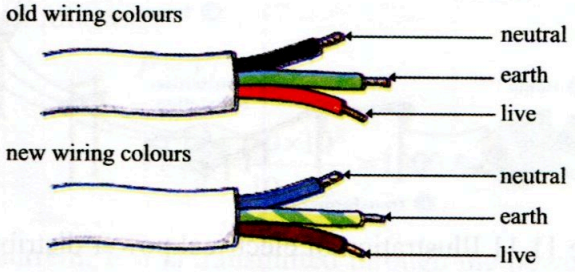


Figure 11.13 The colour coding of electrical wires.

1. Lighting Circuit

In the lighting circuit, the light bulbs are arranged in parallel across the mains line as shown in Figure 11.14. Each bulb has a potential difference of 240 V when connected to the mains line of 240 V. Since the bulbs are connected in parallel, the other bulbs are not affected when one bulb is switched ON or OFF. This means that the lighted bulbs continue to be ON no matter what is done the other bulbs. If the bulbs are connected in series switching OFF one bulb would switch OFF all the others. The lighting circuit differs from the other two in that there are only two wires, live and neutral. A lighting circuit must be connected sequentially as live wire, fuse, switch, load and then to neutral wire.

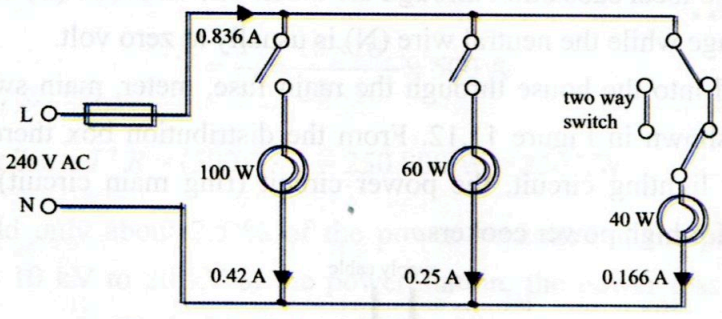


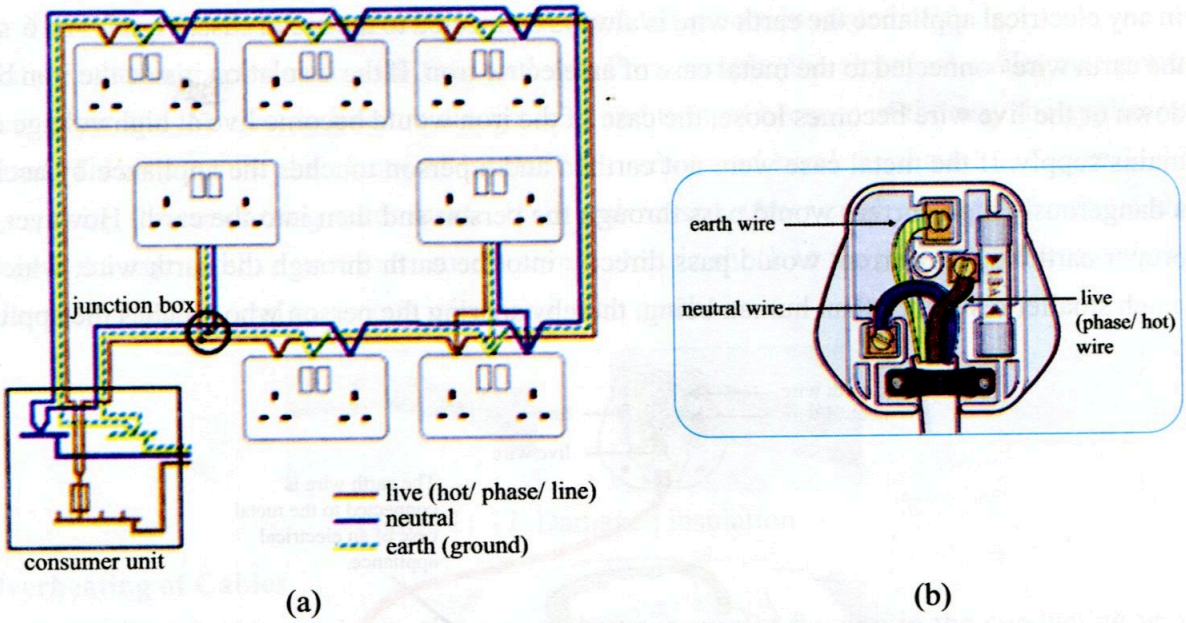
Figure 11.14 A typical lighting circuit

2. Power Circuit (Ring Main Circuit)

A power circuit is an electrical system in which distribution points are connected to the main supply in a continuous closed circuit. The power sockets in a house are connected by means of a ring main circuit. In a ring main circuit the live, neutral and earth wires form a loop of cable going from the consumer unit to all of the sockets in turn and then back to the consumer unit.

Power circuit is for the use of radios, computers, televisions, table lamps and other appliances that consume less power than heating appliances. This circuit supplies electricity to all the wall sockets in the house. Since the current can flow to any particular socket by two ways, thinner wires can be used in the whole ring. Besides both the live and neutral wires running a complete

ring round the house, an earth wire (E) is also added for safety reasons. The power needed to domestic appliances can be taken off any part of the ring. Each main ring circuit also use a 30 A fuse. Three-pin sockets and three-pin plugs are used in the ring main circuit as shown in Figure 11.15.



(a) (b)

Figure 11.15 (a) A ring main circuit with three-pin sockets (b) three-pin plug

3. Heating Circuit

Heating circuit is for the utilization of appliances such as electric cookers, ovens, hot plates, washing machines and dryers. Appliances connected in heating circuits use relatively large amounts of power so that they have their own circuits. Each fuse used in such a circuit is rated at 30 A.

Whatever the type of circuit uses, care must be taken not to be overloaded. And the fuse of each circuit must be connected in the live (L) lead and must never be in the neutral (N) lead. A fuse connected in the neutral lead is pointless; it will in no way protect the appliances used in the circuit.

4. Circuit Breaker

From the consumer unit or distribution box, the wires branch into several parallel circuits for the lights, cookers, immersion heater, and mains sockets. Each circuit passes through a fuse or a circuit breaker. A circuit breaker is an automatic switch which trips (turns OFF) when the current rises above the specified value. It can be reset by turning the switch ON (or) pressing a button.

5. Earthing

Earthing of an electrical appliance is the most important safety precaution. An earthing system (or) grounding system connects specific parts of an electric power system with the ground, typically the earth's conductive surface, for safety and functional purposes.

In any electrical appliance the earth wire is always connected to the metal case. Figure 11.16 shows the earth wire connected to the metal case of an electric iron. If the insulation inside the iron breaks down or the live wire becomes loose, the case of the iron would become live at high voltage of the mains supply. If the metal case were not earthed and a person touches the appliance by accident, a dangerously high current would pass through the person and then into the earth. However, with proper earthing, the current would pass directly into the earth through the earth wire, which has much smaller resistance than human being, thereby sparing the person who touches the appliance.

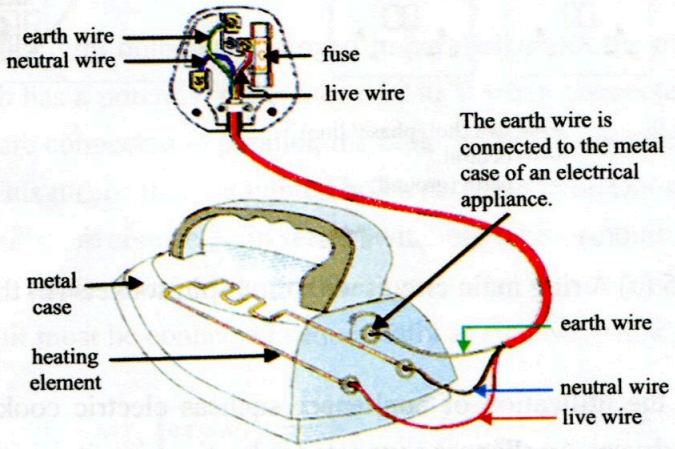


Figure 11.16 Earthing connection of an electric iron

Reviewed Exercise

1. Why are the lamps connected in parallel in the lighting circuit?
2. Why are the fuses and switches placed on the live wire of the circuit?
3. Mention the appropriate wiring circuit for the following electrical appliances (i) bulb and fluorescent lamp (ii) electric iron (iii) water pump motor (iv) electric stove.

Key Words: lighting circuit, ring main circuit, heating circuit, circuit breaker, distribution box, earthing

11.7 DANGERS OF ELECTRICITY

Although electricity plays a very important part in our daily life, it can also be very dangerous if improperly used. Electrical faults in appliances or circuits can cause fires, electric shocks, as well as electrocution to users. Either high current or high voltage can be dangerous to human. The dangers of electricity can be traced to three possible causes: damaged insulation, overheating of cables and damp conditions.

1. Damaged Insulation

All electrical appliances require two wires, the live and neutral wires to form a complete circuit. The live wire is the dangerous wire because it carries a high voltage while the neutral wire is close to zero voltage. These two wires are insulated in vulcanised rubber and they are housed together in either a circular PVC sheathed cable or a circular braided rubber-insulated cable.

Insulating materials deteriorate with time and use. For example, the electrical cables connecting the hair dryer and the electric iron are always bent and twisted because of the way the appliances are used. This will cause the electrical insulation to crack and break as shown in Figure 11.17, thus exposing the conducting wires inside. If the vulcanised rubber covering the live wire is also damaged, the exposed live wire can cause a severe electric shock to the user if the user touches it accidentally. This can lead to serious injury and even death.

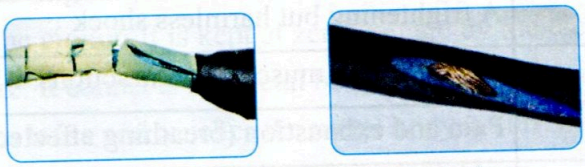


Figure 11.17 Damaged insulation

2. Overheating of Cables

The overheating of cables is due to the extremely large current flowing in the conducting wires under certain conditions such as a short circuit or overloading. A short circuit can result when the live wire makes electrical contact with the neutral wire due to damaged insulation between them. This will produce a large current and the large amount of heat generated can melt the insulation, and start a fire.

Overuse of extension cords and multiple plug adapters on the same circuit are typical causes of an electrical overload by placing too much current on the circuit. Overload also occurs when an equipment draws an excess current from the supply which leads to overheating of the cables. Figure 11.18 shows a typical example of overloading by using too many plugs connected to the same power point.

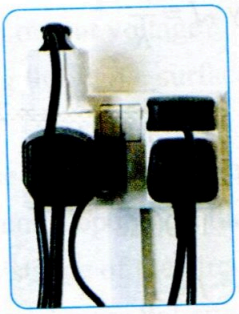


Figure 11.18 Many plugs connected to the same power point

3. Damp Conditions

Many electrical accidents occur in damp conditions in a wet bathroom. In the event of damaged insulation or for any other reason that the live wire should be exposed, the person taking bath

is exposed to the danger of electrocution. If water from the wet floor touches the live wire, it provides a conducting path for a large amount of current to flow through it and through the body of the person in the bath tub. The human body can only withstand up to about 50 mA but the current in this case is very much higher due to the sharp decrease in the resistance of the wet body. Other possible hazards include using hair dryers, electric irons or changing a light bulb with wet hands. Table 11.2 gives the possible physiological effects of an electric current passing through the body.

Table 11.2 Physiological effects of an electric current passing through the body

Current (approximately) (mA)	Effect
1	Threshold (no pain below this point)
5	A frightening but harmless shock
10 – 20	Uncontrolled muscular contractions
50	Pain and exhaustion (breathing affected)
100 – 300	Uncoordinated contraction of the heart leading to death

Since water can conduct electric current, one should never operate an electrical appliance with wet hands. Switches, plugs, sockets, connecting wires, etc. should always be in a dry condition.

Example 11.7 If an electrician accidentally touches a live wire of 220 V, what possible effect would he experience if his skin (i) were dry and had a resistance of 100 kΩ? (ii) were wet and had a resistance of 1000 Ω ?

$V = 220$ V

(i) If his skin were dry, resistance $R = 100\ \text{k}\Omega = 100 \times 10^3\ \Omega$

Current flow through the body $I = \dfrac{V}{R}$

$$= \frac{220}{100 \times 10^3} = 2.2\ \text{mA}$$

The electrician would experience a mild shock.

(ii) If his skin were wet , resistance $R = 1000\ \Omega$

Current flow through the body $I = \dfrac{V}{R}$

$$= \frac{220}{1000} = 220\ \text{mA}$$

The electrician would be electrocuted.

Reviewed Exercise

1. Which quantity determines whether you receive an electrical shock?
2. What causes by the overheating of cable?

Key Words: damaged insulation, overheating, damp conditions, electrocution

SUMMARY

A current flows through a loop of wire whenever there is a change in the magnetic lines of force through the loop. This phenomenon is called **electromagnetic induction**.

The voltage of **live wire** goes alternately negative and positive, making the current flow backwards and forwards through the circuit.

Neutral wire completes the circuit. It is kept at zero voltage by the electricity supply company.

Earth wire is a safety wire. It connects the metal body of the kettle to earth and stops it becoming live. If the live wire comes loose and touches the metal body, a current immediately flows to earth and blows the fuse. This means that the kettle is then safe to touch.

In a transformer, the changing magnetic field not only induces currents in the secondary coil, but also currents in the iron core itself. These currents flow in little circles in the iron core, and are called **eddy currents**.

EXERCISES

1. What is a transformer? How does a step-up transformer differ from a step-down transformer?
2. (i) What is electromagnetic induction? (ii) State the methods by which an induced emf can be produced in a coil.
3. Suggest any two ways to increase the output voltage of an AC generator.
4. In an AC generator, the speed at which the coil rotates is tripled. How would this affect (i) the frequency (ii) the maximum output voltage?
5. Explain how magnetic flux passing through a surface can be zero when the magnetic field is not zero.
6. Discuss the factors determining the induced emf in a closed loop of wire.
7. (i) Does the induced emf in a circuit depend on the resistance of the circuit? (ii) Does the induced current depend on the resistance of the circuit?
8. Why are ordinary circuit breakers and fuses not effective in preventing shocks?
9. A square loop of wire of side 5.0 cm is in a uniform magnetic field 0.16 T. What is the magnetic flux in the loop (i) when magnetic field is perpendicular to the surface of the loop and (ii) when magnetic field is at an angle of 30° to the surface of the loop?

10. A circular coil of radius 8 cm and 20 turns is rotated about its vertical diameter with an angular speed of 150 rad s^{-1} in a uniform horizontal magnetic field of 3×10^{-2} T. Obtain the maximum emf induced in the coil. If the coil forms a closed loop of resistance 10 Ω, calculate the maximum value of current in the coil.
11. The armature of a 60 Hz AC generator rotates in a magnetic field of 0.15 T. If the area of the coil is 2.0×10^{-2} m^2, how many loops must the coil contain if the maximum output is to be 170 V?
12. One coil of a transformer has 50 turns and the other coil has 1000 turns. The alternating voltage of 220 V is applied to the primary. Find the alternating voltage from the secondary when it is used as (i) a step-up transformer and (ii) a step-down transformer.
13. The primary coil of a step-up transformer having 220 turns is connected to a 110 V alternating voltage source. The secondary delivers a voltage of 12 000 V and a current of 40 mA. (i) How many turns are there in the secondary? (ii) What is the current in the primary? (iii) What is the power output?
14. An electric welding machine uses a current of 400 A. The machine has a transformer whose primary coil has 400 turns and draws 4 A from a 220 V power line. (i) How many turns does the secondary coil have? (ii) What is the voltage across the secondary coil?
15. A transformer has 2000 turns in primary, 100 turns in secondary and its efficiency 100 %. It is used for 10 V, 40 W bulb. Calculate (i) the supply voltage and (ii) the current through the input coil.
16. 120 kW of electric power is sent to a small town from a power plant 10 km away. The transmission lines have a resistance of 0.40 Ω per km. Calculate the power loss if the power is transmitted at (i) 240 V and (ii) 24 000 V.

CONCEPT MAP

ELECTROMAGNETIC INDUCTION, GENERATION AND DISTRIBUTION OF ELECTRICITY

Electromagnetic Induction
- magnetic flux density
$B = \mu H$
- magnetic flux
$\phi_B = \int \vec{B} \cdot d\vec{A}$

Generation and Distribution of Electricity
- alternating emf
$E = E_m \sin \omega t$
- alternating current
$I = I_m \sin \omega t$

Faraday's law and Lenz's law

$$E = -N \frac{d\phi_B}{dt}$$

Applications of electromagnetic induction
- induction coil
- AC generator

$E = N B A \omega \sin \omega t$

- transformer

$$\frac{V_s}{V_p} = \frac{E_s}{E_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s}$$

Power transmission
- grid system

Principles of house wiring
- lighting circuit
- power circuit
- heating circuit
- circuit breaker
- earthing

Dangers of electricity
- damaged insulation
- overheating of cables
- damp conditions

CHAPTER 12

DIGITAL ELECTRONICS AND COMMUNICATION SYSTEM

The world's reliance on electronics is so great that commentators claim people live in an electronic age. Nowadays, people are surrounded by electronic appliances such as televisions, radios, computers, mobile phones, laptops, and DVD players; and also the products of electric components such as hearing aids, pacemakers, blood pressure monitor, oximeter and various medical instruments, microwave ovens, refrigerators, washing machines and other kitchen appliances, automatic vehicles, robotics, as well as numerous applications in diverse areas.

Learning Outcomes

It is expected that students will

- develop conceptual understanding of the analogue and digital electrical/electronic circuits, and aware of new and emerging technologies.
- describe how light emitting diodes (LEDs) and light dependent resistor (LDR) work and how they are made.
- understand the structure of liquid crystal displays (LCDs).
- identify common logic gates, their functions and truth tables.
- determine the corresponding Boolean expression for the circuit output.
- understand the term modulation and be able to distinguish between amplitude modulation (AM) and frequency modulation (FM).
- explore the technology and operation of radio receiver.

12.1 ELECTRONIC SYSTEM

The branch of physics and technology concerned with the design of circuits using transistors, integrated circuits and microchips, and also with the behaviour and movement of electrons in a semiconductor, conductor or vacuum is known as electronics.

Electronic devices contain circuitry that involve active electrical components and associated passive interconnection technologies. Active electrical components are vacuum tubes, transistors, diodes, ICs (Integrated Circuits), TRIACs (Three terminal AC switches), SCRs (Silicon Controlled Rectifiers), LEDs (Light Emitting Diodes), etc. and requires a power source to operate. Passive electrical components are capable of operating without an external power source. Passive electrical components are resistors, capacitors, inductors, transformers, LDR, transducer, etc.

In Grade 11, we have already mentioned electronic devices which are fabricated using semiconductors: pn junction diode and transistor. Now, we are going to introduce some special

applications of junction diode such as zener diode and optoelectronic devices such as LEDs, and photodiode.

1. Zener Diode

Zener diode is a heavily doped semiconductor (silicon or germanium) pn junction diode. It makes the current flow in the backward direction when reverse biased. It operates like the normal diode when in the forward-bias mode, and has a turn-on voltage of between 0.3 and 0.7 V.

When the voltage across the terminals of a zener diode is reversed and the potential reaches the zener voltage (V_Z) (breakdown voltage or knee voltage), the junction breaks down and the current flows in the reverse direction. This effect is known as the zener effect. That is, as the reverse voltage increases to the breakdown voltage, large current starts flowing through the diode. The symbol and characteristics of a zener diode is shown in Figure 12.1.

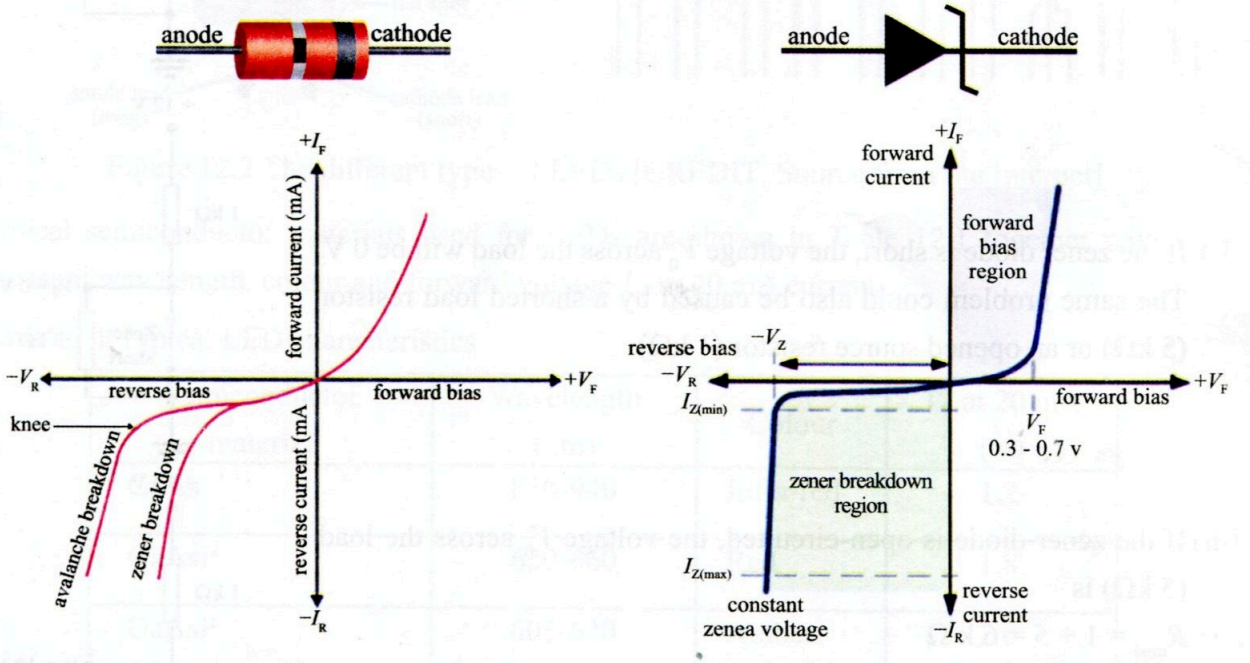


Figure 12.1 The physical, circuit symbol and characteristics of zener diode

The main difference between a zener diode and a normal diode lies in the passage of current. A normal diode allows current to flow only in one direction while zener diode allows current to flow in both directions.

There are two types of reverse breakdown regions in a zener diode: (i) avalanche breakdown and (ii) zener breakdown.

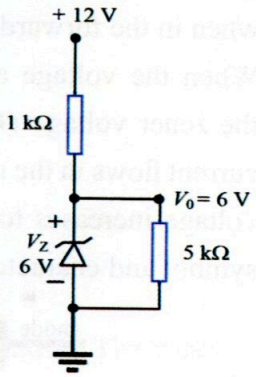
Avalanche breakdown occurs at junctions which being lightly doped have wide depletion layers.

Zener breakdown occurs at junctions which being heavily doped have narrow depletion layers.

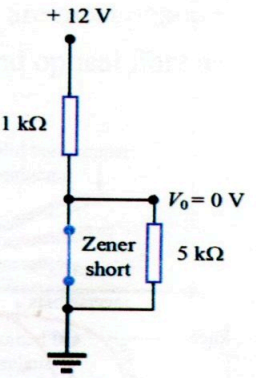
For normal operation of a zener diode, in breakdown region, the current through the diode should be limited by an external circuit. It has a reverse-breakdown voltage at which the diode starts to

conduct electric current, and remains continuous in the reverse-bias mode. The voltage drop across the zener diode always remains constant irrespective of the applied voltage, and this feature of the zener diode makes it suitable for voltage regulation as reference elements, surge suppressors, and in switching applications and clipper circuits.

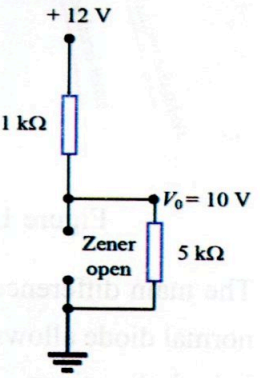
Example 12.1 Figure shows the basic zener diode circuits. What will be the circuit behaviour if the zener is (i) working properly (ii) shorted and (iii) open-circuited?



(i) If the zener diode is working properly, the voltage V_0 across the load (5 kΩ) will be nearly 6 V.



(ii) If the zener diode is short, the voltage V_0 across the load will be 0 V. The same problem could also be caused by a shorted load resistor (5 kΩ) or an opened source resistor (1 kΩ).



(iii) If the zener diode is open-circuited, the voltage V_0 across the load (5 kΩ) is

$$R_{total} = 1 + 5 = 6\ \text{k}\Omega$$

$$I = \frac{V}{R_{total}} = \frac{12}{6} = 2\ \text{mA}$$

$$V_{5\,\text{k}\Omega} = I R_L = 2 \times 5 = 10\ \text{V}$$

2. Light Emitting Diode (LED)

A LED is a semiconductor light source that emits light when current flows through it. A light emitting diode (LED) is known to be one of the best optoelectronic devices. It is a semiconductor light source capable of emitting a fairly narrow bandwidth of visible (or) invisible light when its internal diode junction attains a forward-biased current (or) voltage. The visible lights that an LED emits are usually orange, red, yellow, or green. The commonly used invisible light for LED is infrared light. A modern light source that requires low voltages is the light emitting diode (LED).

The LEDs are made from exotic semiconductor compounds such as Gallium Arsenide (GaAs), Gallium Phosphide (GaP), Gallium Arsenide Phosphide (GaAsP), Silicon Carbide (SiC) or Gallium Indium Nitride (GaInN) all mixed together at different ratios to produce a distinct wavelength of colour. In LED the energy is radiated in the form of light. LEDs do not directly produce white light. Using a blue LED with a phosphor coating to convert blue light to white light by a process called fluorescence. Different types of LEDs are shown in Figure 12.2.

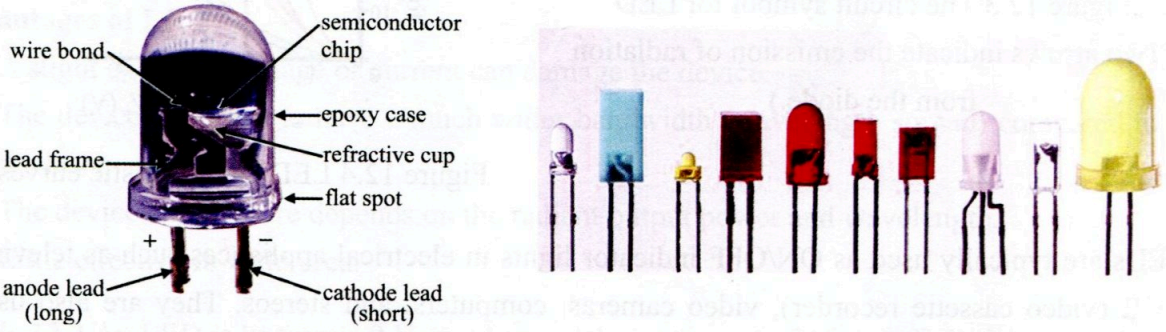


Figure 12.2 The different types of LEDs [CREDIT: Source from the Internet]

Typical semiconductor materials used for LEDs are shown in Table 12.1 together with their emission wavelength, colour and forward voltage V_F at 20 mA current.

Table 12.1 Typical LED characteristics

Semiconductor material	Wavelength (nm)	Colour	V_F at 20 mA (V)
GaAs	850-940	Infra-red	1.2
GaAsP	630-660	Red	1.8
GaAsP	605-620	Amber	2.0
GaAsP:N	585-595	Yellow	2.2
AlGaP	550-570	Green	3.5
SiC	430-505	Blue	3.6
GaInN:P		White	4.0

The circuit symbol of LED and the forward biased LED characteristic curves are shown in Figure 12.3 and 12.4.

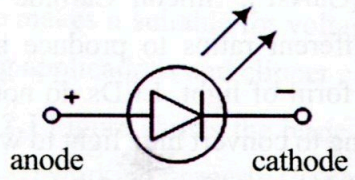


Figure 12.3 The circuit symbol for LED
(Two arrows indicate the emission of radiation from the diode.)

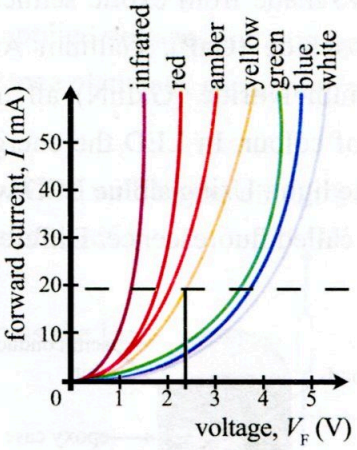


Figure 12.4 LED characteristic curves

LEDs are typically used as ON/OFF indicator lights in electrical appliances such as televisions, VCR (video cassette recorder), video cameras, computers, and stereos. They are also used to display numbers in some alarm clocks, radios, and microwave ovens. Another use is very large video displays at sporting events and concerts.

LED as an Indicator

Figure 12.5 is an LED indicator circuit, one of the main applications of LED. A current limiting resistor should be connected in series with LED in the circuit to protect the LED. The value of the series resistance should be chosen that the voltage drop at LED must be nearly equal to that at the limiting resistor.

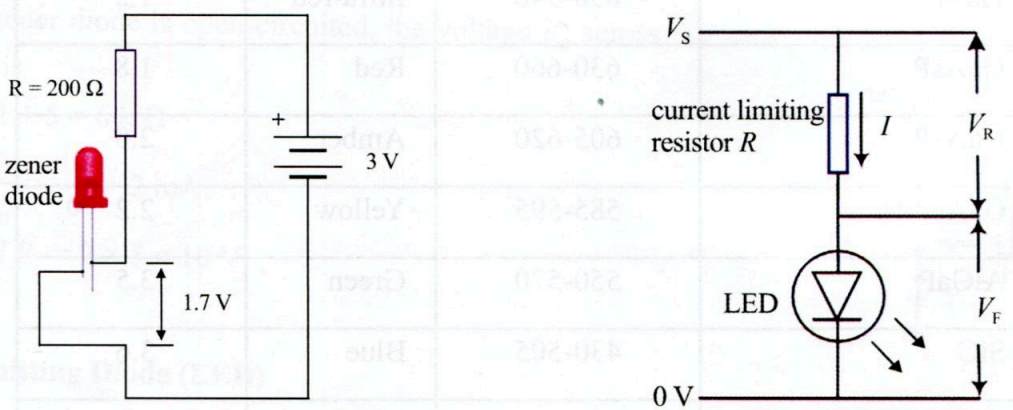


Figure 12.5 Simple LED indicator circuit

In Figure 12.5, let V_S be the supply voltage, V_F be the required forward bias voltage for the LED, I be the desired current flowing through LED. Then, the current limiting resistance R is

$$V_R = V_S - V_F$$

$$R = \frac{V_R}{I} = \frac{V_S - V_F}{I}$$

Advantages of LED

- Very low voltage (1 to 2 V) and current (5 to 20 mA) are enough to drive the LED.
- Less power consumption (less than 150 mW).
- Fast action and no warmup time. (The response time is about 10 ns)
- Miniature in size and hence light weight.
- Long life span and ruggedness.
- More reliable, more efficient to use under cold temperature.

Disadvantages of LED

- A slight excess in voltage or current can damage the device.
- The device is known to have a much wider bandwidth (wavelength spread) compared to the laser.
- The device temperature depends on the radiant output power and wavelength.
- Little effective in wide-area.

Example 12.2 An LED is lit from a 9 V supply and takes a current of 15 mA. Calculate the value of the current limiting resistor required for the LED. (Assume that the forward voltage drop will be 2 V)

$V_s = 9$ V, $V_F = 2$ V, $I = 15$ mA $= 15 \times 10^{-3}$ A

The current limiting resistor required for the LED,

$$R = \frac{V_R}{I} = \frac{V_S - V_F}{I}$$

$$= \frac{9-2}{15 \times 10^{-3}} = 466.67\ \Omega \text{ (in practice, 470 } \Omega \text{ is used)}$$

Reviewed Exercise

1. Where are LEDs used?
2. Describe some applications of LEDs.
3. An amber-coloured LED with a forward voltage drop of 2 V is to be connected to a 5 V stabilized DC power supply. Calculate the value of the series resistor required to limit the forward current to less than 10 mA. Also calculate the current flowing through the diode if a 100 Ω of current limitting resistor is used.

Key Words: zener diode, avalanche breakdown, zener breakdown, light emitting diode, zener effect, forward biased current

3. Photodiode

A photodiode is a semiconductor device that converts light into electrical current. The current is generated when photons are absorbed in the photodiode. Photodiodes work in reverse biased mode. Figure 12.6 shows photodiode and its circuit symbol.

Figure 12.6 Photodiode and its circuit symbol

When the photodiode is illuminated with light (photons) of energy $h\nu$ greater than the energy gap E_g of the semiconductor, then electron-hole pairs are generated due to absorption of photons. When an external load is connected, the current flows. This current is called photocurrent. The magnitude of the photocurrent depends on the intensity of incident light. Thus, the photodiode can be used as a photodetector to detect the optical signals.

A small amount of current (dark current), due to the flow of minority carriers, is also produced when no light is falling. Silicon is most often used to fabricate photodiodes, though germanium and gallium arsenide can also be used.

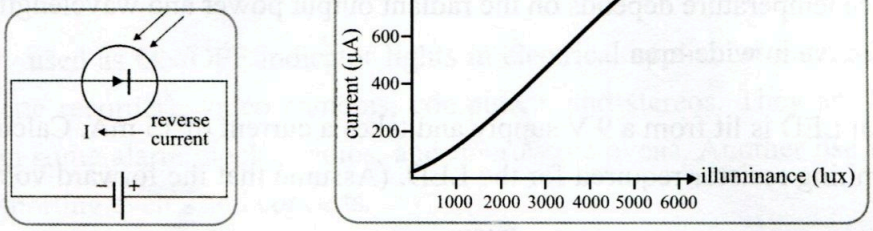


Figure 12.7 Simple circuit diagram of photodiode and its characteristics curve

Photodiodes are used as photosensors in compact disc (CD) players, televisions, air conditioners and many other electronics devices. Photodiodes can also be used for light measurement in light meters and to respond light levels, and in switching on street lighting after dark.

Photocurrent is an electric current produced by electromagnetic radiation in the photoelctric effect, photovoltaic effect or photoconductivity.

Reviewed Exercise

- Explain symbol, circuit diagram and function of a photodiode.

Key Words: photodiode, energy gap, photosensors, electron-hole pairs, photocurrent, photon, dark current

4. Light Dependent Resistor (Photoresistor/ Photocell)

A light dependent resistor (LDR) is a light controlled variable resistor. LDR consists of a disc of semiconductor material on its surface with two electrodes. LDR is also called light sensitive variable resistors. Its resistance depends on the intensity of light falling on its surface. In the dark or in dim light, the material of the disc has a relatively small number of free electrons in it. That is few electrons to carry electric charge. It becomes a poor conductor of electric current. Its resistance is quite high (~ MΩ) called dark resistance. In the bright light, more electrons escape from the atoms of the semiconductor. Since there are more electrons to carry electric charge, it becomes a good conductor. Its resistance is low (few hundred Ω). As light level increases, the LDR resistance decreases. Figure 12.8 shows LDR and its circuit symbol.

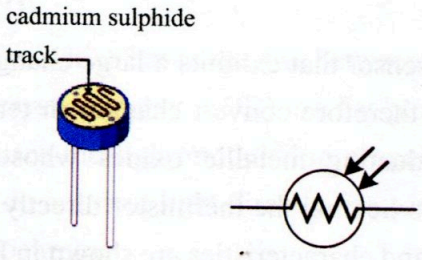


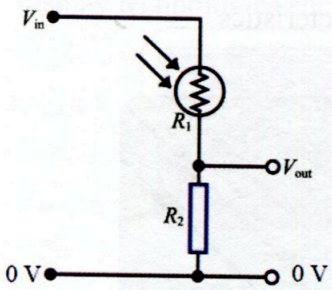
Figure 12.8 Light dependent resistor and its circuit symbol

When LDR is exposed to light, it takes a few milliseconds to lowers its resistance and it may take a few seconds to return to its dark resistance after removal of light.

Special semiconductor crystal, such as cadmium sulfide (or) lead sulfide is used to make photo resistors. LDRs are used as light activated relay switches (to trip a relay whenever the light intensity changes) in burglar alarm, alarm clock, light intensity meters, etc.

Example 12.3 In the figure the resistance of R_2 is 2000 Ω and V_{in} is 6 V. (i) In daylight the resistance of R_1 is 500 Ω. Calculate the voltage across R_2. (ii) In the dark the resistance of R_1 is 198 000 Ω. Calculate the voltage across R_2.

$R_2 = 2000\ \Omega$, $V_{in} = 6$ V



(i) In daylight, $R_1 = 500\ \Omega$

Since $I = \dfrac{V_{in}}{(R_1 + R_2)}$ and $V_{out} = I\,R_2$

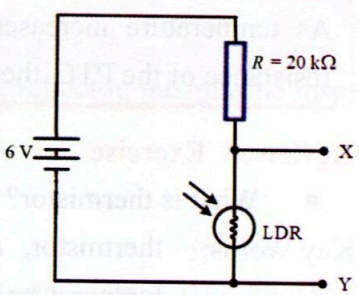
$$V_{out} = V_{in} \times \frac{R_2}{(R_1 + R_2)} = 6 \times \frac{2000}{(500 + 2000)} = 4.8 \text{ V}$$

(ii) In dark, $R_1 = 198000\ \Omega$

$$V_{out} = V_{in} \times \frac{R_2}{(R_1 + R_2)} = 6 \times \frac{2000}{(198000 + 2000)} = 0.06 \text{ V}$$

Reviewed Exercise

1. What effect does an increase of light intensity on LDR have on its resistance?
2. The circuit shows a potential divider, which consists of a fixed resistor and a light dependent resistor (LDR). The potential divider is used to switch on a lamp when it gets dark. The resistance of fixed resistor is 20 kΩ. The potential difference across XY is 4 V when the lamp is switched on. What is the resistance of the light-dependent resistor (LDR)?



(Hint: $V_S = V_R + V_{LDR}$, $I = \dfrac{V_R}{R}$ and $R_{LDR} = \dfrac{V_{LDR}}{I}$)

Key Words: light dependent resistor, photoresistors, activated relay switch, dark resistance

5. Thermistor

A thermistor is a temperature sensor that exhibits a large change in resistance proportional to a change in temperature. It can therefore convert changes in temperature into changes in electric current. It contains semiconducting metallic oxides whose resistance decreases when the temperature raises either due to heating the thermistor directly or to passing a current through it. Thermistor, its circuit symbol and characteristics are shown in Figure 12.9. There are two types of thermistors. The most commonly used is the Negative Temperature Coefficient (NTC) thermistor. The NTC's resistance decreases as the temperature increases, and vice versa. In the Positive Temperature Coefficient (PTC) thermistor, the resistance increases as the temperature increases, and is used to break the circuit at preset temperatures. The various shapes of thermistors are as shown in Figure 12.10.

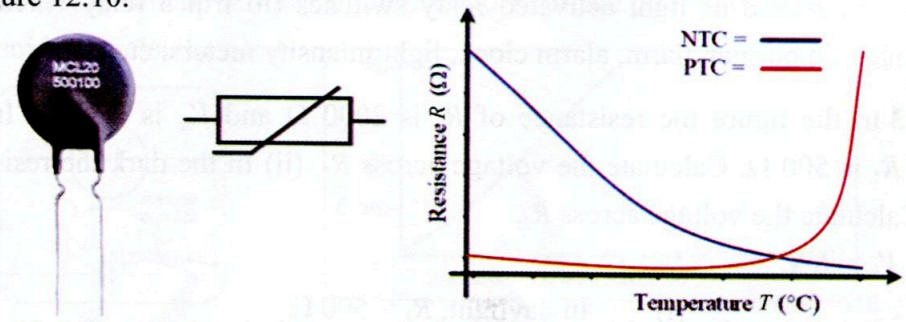


Figure 12.9 Thermistor, circuit symbol and its characteristics

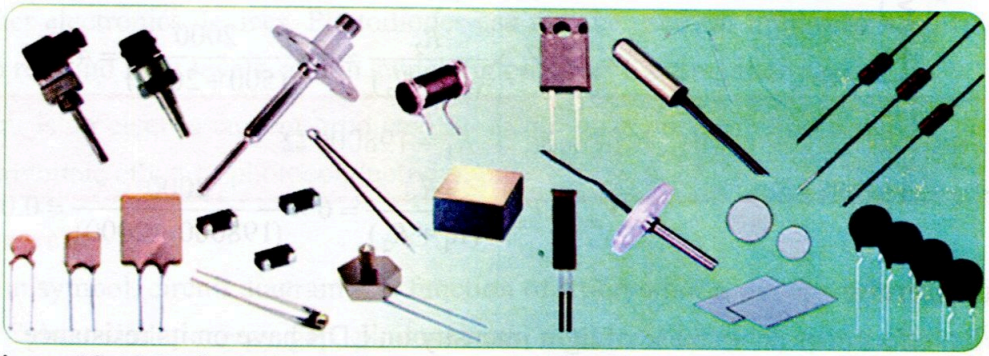


Figure 12.10 Various shapes of thermistor [CREDIT: Source from the Internet]

As temperature increases, the resistance of the NTC thermistor decreases whereas the resistance of the PTC thermistor increases.

Reviewed Exercise

- What is thermistor? What is the main difference between NTC and PTC thermistor?

Key Words: thermistor, negative temperature coefficient, positive temperature coefficient, temperature sensor

6. Liquid Crystal Display (LCD)

Liquid crystals are a state of matter which has properties between those of conventional liquids and those of solid crystals. A liquid crystal display (LCD) is a flat-panel display or other electronically modulated optical device that uses the light-modulating properties of liquid crystals combined with polarizers. Liquid crystals do not emit light directly, instead using a backlight or reflector to produce images in color or monochrome.

LCDs are lit by a backlight, and pixels are switched ON and OFF electronically while using liquid crystals to rotate polarized light. A polarizing glass filter is placed in front and behind all the pixels, the front filter is placed at 90 degrees. In between both filters are the liquid crystals, which can be electronically switched ON and OFF.

LCDs can be commonly found in smartphones, televisions, computer monitors and instrument panels (Figure 12.11). It is a big leap in technology in which LCDs replace light-emitting diodes (LEDs) and gas-plasma displays. LCDs allowed displays to be much thinner than cathode ray tube (CRT) technology. LCDs consume much less power than LEDs and gas-plasma displays because they work on the principle of blocking light rather than emitting it. Where an LED emits light, the liquid crystals in an LCD produce an image using a backlight. Although LCDs have replaced older display technologies (i.e. CRT), LCDs have begun being replaced by new display technologies such as organic light-emitting diodes (OLEDs).

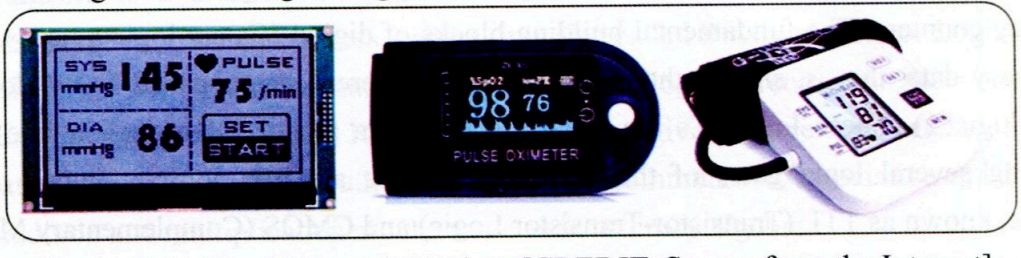


Figure 12.11 Liquid Crystal Displays [CREDIT: Source from the Internet]

A polarizer is an optical device that can convert an unpolarized light wave into a polarize light wave by blocking all other vibrations.

Reviewed Exercise

- What is the disadvantage of LCD displays?

Key Words: liquid crystal display, organice light emitding diode, reflector, gas-plasma display, polarizer

12.2 DIGITAL ELECTRONICS

Digital electronics is a field of electronics involving the study of digital signals and the engineering of devices that use or produce them. This is in contrast to analog electronics and analog signals.

Digital electronic circuits are usually made from large assemblies of logic gates, often packaged in integrated circuits. Complex devices may have simple electronic representations of Boolean logic functions. An advantage of digital circuits when compared to analog circuits is that signals represented digitally can be transmitted without degradation caused by noise.

Many of our household items make use of digital electronics such as computers, laptops, televisions, remote controls and other entertainment systems (or) kitchen appliances like dishwashers and washing machines.

An IC (integrated circuit), sometimes called a chip (or) microchip, is a semiconductor wafer on which thousands (or) millions of tiny resistors, capacitors and transistors are fabricated. An IC can function as an amplifier, oscillator, timer, counter, computer memory, or microprocessor. A particular IC is categorized as either linear analog (or) digital, depending on its intended application. ICs are of linear, digital and mixed types.

Linear ICs have continuously variable output (theoretically capable of attaining an infinite number of states) that depends on the input signal level. As the term implies, the output signal level is a linear function of the input signal level. Linear ICs are used as af (audio-frequency) and rf (radio-frequency) amplifiers.

Digital ICs operate at only a few defined levels or states, rather than over a continuous range of signal amplitudes. These devices are used in computers, computer networks, modems and frequency counters. The fundamental building blocks of digital ICs are logic gates, which work with binary data, that is, signals that have only two different states, called LOW (logic 0) and HIGH (logic 1). These logic ICs are usually supplied in plastic DIL (dual in line) packages containing several logic gates of the same type. There are two common types of package available, known as TTL (Transistor-Transistor Logic) and CMOS (Complementary Metal Oxide Semiconductor).

Boolean Algebra and Logic Gates

In Grade 11, we have described the five common logic gates. Now we will give a brief review of these gates. The symbols and truth tables for the different gates are given in Figure 12.12 and Table 12.2. The AND gate with inverted output is called a NAND gate. The OR gate with inverted output is called a NOR gate. The inverted connection (inverted gate) is indicated by a bubble.

There are two remaining gates of the primary electronics logic gates: XOR, which stands for Exclusive OR, and XNOR, which stands for Exclusive NOR. In an XOR gate, the output is HIGH if one, and only one, of the inputs is HIGH. That is, XOR gate produces a 0 when both inputs match. The XOR gate operator is written as $A \oplus B$. An XNOR gate is an XOR gate whose output is inverted. Thus, XNOR gate operator is written as $\overline{A \oplus B}$.

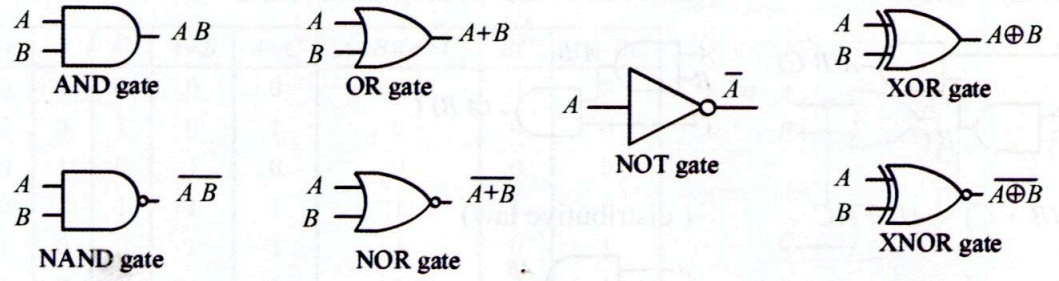


Figure 12.12 Gate symbols

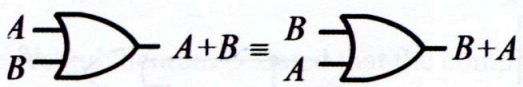
Table 12.2 The truth table for different gates

A	B	NOT $\bar{A}$	AND AB	NAND $\overline{AB}$	OR $A+B$	NOR $\overline{A+B}$	XOR $A \oplus B$	XNOR $\overline{A \oplus B}$
0	0	1	0	1	0	1	0	1
0	1	1	0	1	1	0	1	0
1	0	0	0	1	1	0	1	0
1	1	0	1	0	1	0	0	1

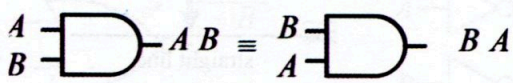
The Basic Laws of Boolean Algebra and De Morgan's Theorems

Boolean algebra is a branch of mathematics where variables can have only two possible values: false and true, (or) logic values 0 and 1. It is also called Binary algebra or Logical algebra. Boolean algebra is used to analyze and simplify the digital (logic) circuits. The basic operations in Boolean algebra are AND, OR, and NOT. Boolean algebra gives a more compact way to describe a combinational logic circuit than truth tables alone. The commutative, associative and distributive laws for addition and multiplication are the same as in ordinary algebra.

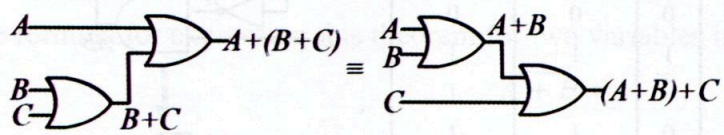
1. $A + B = B + A$ (commutative law of addition)



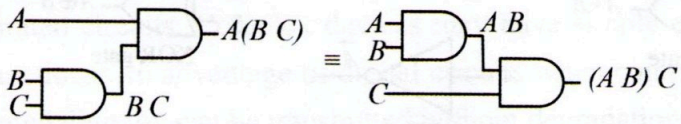
2. $A\ B = B\ A$ (commutative law of multiplication)



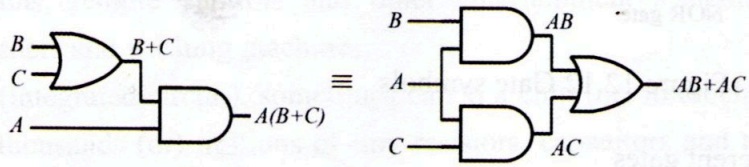
3. $A + (B + C) = (A + B) + C$ (associative law of addition)



4. $A(BC) = (AB)C$ (associative law of multiplication)



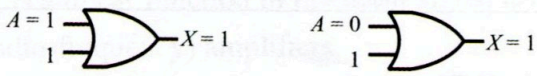
5. $A(B + C) = AB + AC$ (distributive law)



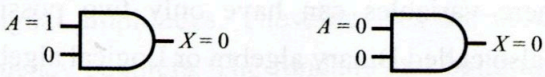
Basic Rules of Boolean Algebra

1. $A + 0 = A$

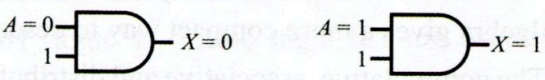
2. $A + 1 = 1$



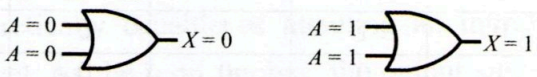
3. $A \cdot 0 = 0$



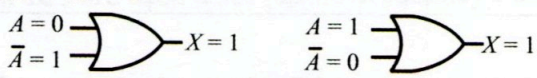
4. $A \cdot 1 = A$



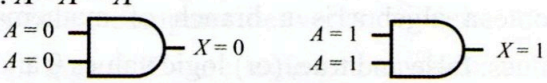
5. $A + A = A$



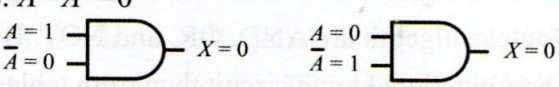
6. $A + \bar{A} = 1$



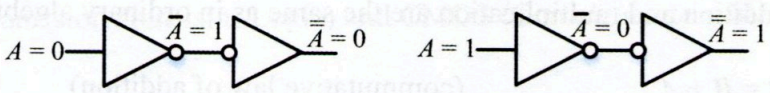
7. $A \cdot A = A$



8. $A \cdot \bar{A} = 0$



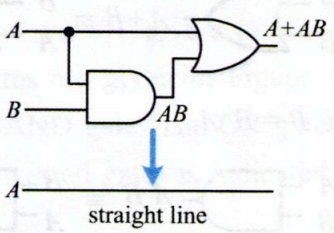
9. $\bar{\bar{A}} = A$



10. $A + AB = A$

A	B	AB	A + AB
0	0	0	0
0	1	0	0
1	0	0	1
1	1	1	1

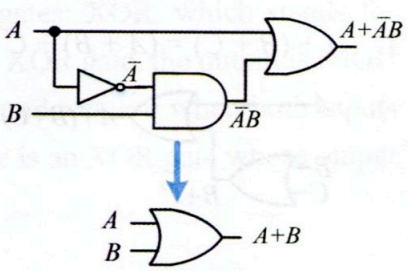
equal



11. $A + \bar{A}B = A + B$

A	B	ĀB	A + ĀB	A + B
0	0	0	0	0
0	1	1	1	1
1	0	0	1	1
1	1	0	1	1

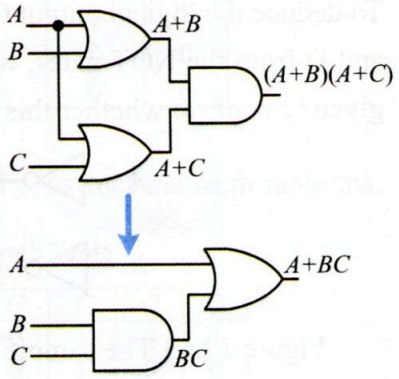
equal



12. $(A + B)(A + C) = A + BC$

A	B	C	$A+B$	$A+C$	$(A+B)(A+C)$	BC	$A + BC$
0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0
0	1	0	1	0	0	0	0
0	1	1	1	1	1	1	1
1	0	0	1	1	1	0	1
1	0	1	1	1	1	0	1
1	1	0	1	1	1	0	1
1	1	1	1	1	1	1	1

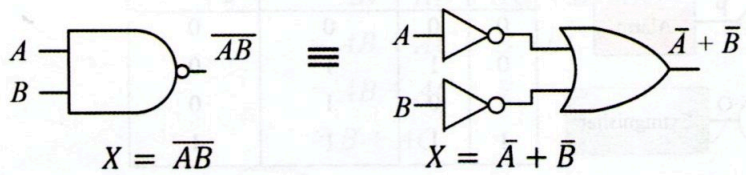
equal



De Morgan's Theorem

Augustus De Morgan formulated an extension to English mathematician George Boole's algebraic logic that has become very important in digital logic. The theorem comprises two laws that describe how inverting the inputs to a gate, changes the gate's function.

The first theorem is stated that the complement of two or more ANDed variables is equivalent to the OR of the complements of the individual variables.

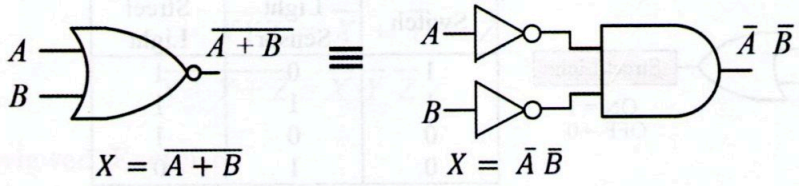


input		output	
A	B	$\overline{AB}$	$\bar{A} + \bar{B}$
0	0	1	1
0	1	1	1
1	0	1	1
1	1	0	0

The formula for expressing this theorem for two variables is

$$\overline{AB} = \bar{A} + \bar{B}$$

The second theorem is stated that the complement of two or more ORed variables is equivalent to the AND of the complements of the individual variables.



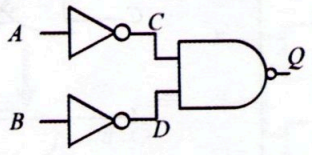
input		output	
A	B	$\overline{A+B}$	$\bar{A}\bar{B}$
0	0	1	1
0	1	0	0
1	0	0	0
1	1	0	0

The formula for expressing this theorem for two variables is

$$\overline{A+B} = \bar{A}\,\bar{B}$$

Combination and Uses of Logic Gates

Consider a system that consists of two NOT gates and one NAND gate as shown in Figure 12.13. To deduce the logical output Q of the system, we have to work out first the intermediate outputs C and D from the NOT gates, which act as the two inputs to the NAND gate. Check the truth table given to confirm whether this system is equivalent to an OR gate.



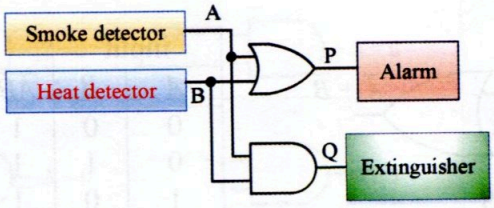
A	*B*	*C*	*D*	*Q*
0	0	1	1	0
0	1	1	0	1
1	0	0	1	1
1	1	0	0	1

Figure 12.13 The sample circuit diagram of deduce the logical output and the truth table

Logic gates can be used as processors in electronic control systems. Many of these can be demonstrated by connecting together commercial modules.

(a) Fire Alarm System

The fire alarm system consists of smoke detector, heat detector and two logic gates as shown in Figure 12.14. The fire alarm will turn on when either smoke or heat is detected. If both smoke and heat are detected, the fire extinguisher will also be set to operate.

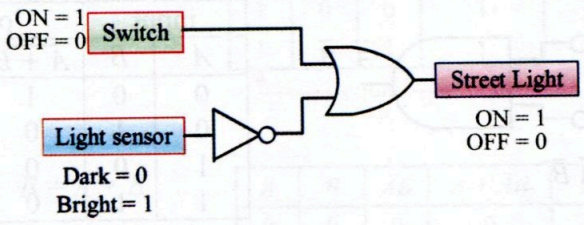


A	*B*	*P* = *A*+*B*	*Q* = *AB*
0	0	0	0
0	1	1	0
1	0	1	0
1	1	1	1

Figure 12.14 The circuit diagram of fire alarm system and its truth table

(b) Street Lighting System

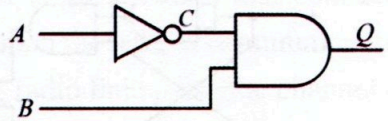
A system is required which allows the street lights either to be turned on manually by a switch at any time or automatically by a light sensor when it is dark. The arrangement in Figure 12.15 achieves this since the OR gate gives a 1 output when either or both of its inputs are 1.



Switch	Light Sensor	Street Light
1	0	1
1	1	1
0	0	1
0	1	0

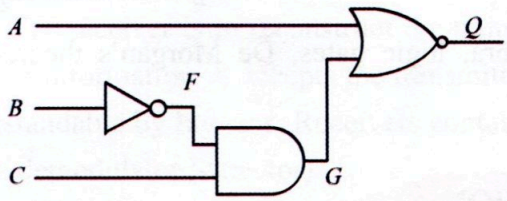
Figure 12.15 The circuit diagram of street lighting system and its truth table

Example 12.3 Check the following logic system carefully and then construct the truth table that follows:



A	B	C	Q
0	0	1	0
0	1	1	1
1	0	0	0
1	1	0	0

Example 12.4 Check the following logic system carefully and then construct the truth table that follows:

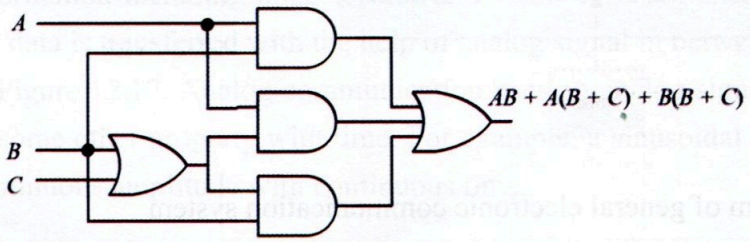


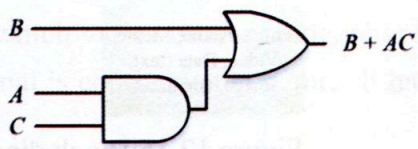
Truth Table

A	B	C	F	G	Q
0	0	0	1	0	1
1	0	0	1	0	0
0	1	0	0	0	1
1	1	0	0	0	0
0	0	1	1	1	0
1	0	1	1	1	0
0	1	1	0	0	1
1	1	1	0	0	0

Example 12.5 Simplify the expression $AB + A(B + C) + B(B + C)$ and draw the equivalent gate diagram.

$$AB + A(B + C) + B(B + C) = AB + AB + AC + BB + BC$$
$$= AB + AB + AC + B + BC \qquad (BB = B)$$
$$= AB + AC + B + BC \qquad (AB + AB = AB)$$
$$= AB + AC + B \qquad (B + BC = B)$$
$$= B + AC \qquad (AB + B = B)$$



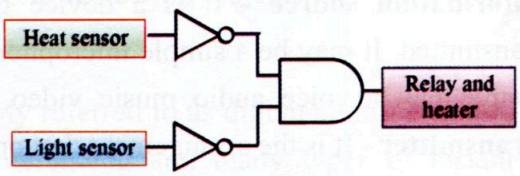


Example 12.6 Apply De Morgan's theorems to the expression $\overline{XYZ}$ and $\overline{X + Y + Z}$.

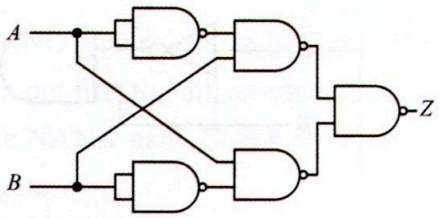
$$\overline{XYZ} = \overline{X} + \overline{Y} + \overline{Z}$$
$$\overline{X + Y + Z} = \overline{X}\,\overline{Y}\,\overline{Z}$$

Reviewed Exercise

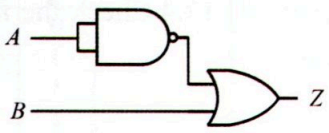
1. State the names of the two types of logic gate used in following figure. Explain what happens to the outputs of the three logic gates when (i) it is cold (0) and dark (0), (ii) it is cold (0) and bright (1).



2. Write the truth table of the following circuit. To what gate is it equivalent?



3. For the circuit below, write the value of Z as a Boolean expression, using A and B.



Key Words: integrated circuits, Boolean algebra, logic gates, De Morgan's theorem, TTL, CMOS

12.3 BASIC ELECTRONIC COMMUNICATION

In general, an electronic communication system is composed of the transmission, reception, and processing of information between two or more locations with the use of electronic circuits. The basic components of an electronic communication system are the transmitter, communication channel or medium receiver and noise (as a source of disturbance). The block diagram shown in Figure 12.16 depicts the general form of a communication system.

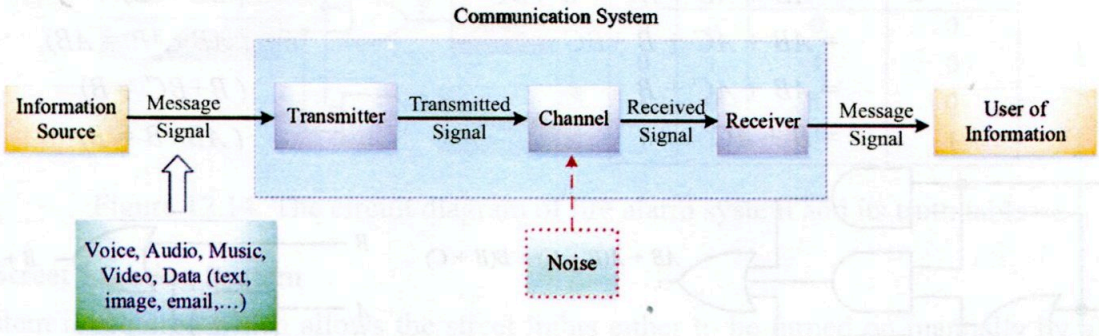


Figure 12.16 Block diagram of general electronic communication system

Analog signals (human voice) or digital signals (binary data) are inputted to the system, processed within the electronic circuits for transmission, then decoded by the receiver. The system is claimed to be reliable and effective only errors are minimized within the process. Examples: internet, public switched telephone network, intranet and extranet, smart home systems and television.

Information source - It is a device that generates information (messages or data) to be transmitted. It may be a simple microphone and computer keyboard. The messages generated by source may be voice, audio, music, video, data (text, image, email, temperature, pressure, etc.)

Transmitter - It is the arrangement that processes the message signal into an appropriate form for transmission and subsequent reception.

Channel - It is the medium that connects transmitter and receiver to convey the information. Depending upon the type of communication system, a channel may be copper wires, coaxial cable, optical fibre, radio links, satellite channel or combination of any of these.

Noise - Noise is an unwanted signal that disturbs, interferes and affects the transmitted signal. Noise cannot be prevented but can be minimized. The measure of noise is usually expressed in terms of SNR (signal to noise ratio). The external noise is often minimized and eliminated by the appropriate design of the channel, shielding of cables. Digital signals are more immune to noise than analog signals.

Receiver - The receiver is to reconstruct the transmitted signal and delivers it to the destination called user of information. It accepts the transmitted message from the channel and converts to a form understandable by humans. Receivers contain amplifiers, oscillators, mixers, tuned circuits, filters and a demodulator (detector).

Modes of Communication System

There are two basic modes of communication: point-to-point and broadcast. In point-to-point mode, communication takes place over a link between a single transmitter and a receiver. Telephony is an example of point-to-point communication. In the broadcast mode, there are a large number of receivers corresponding to a single transmitter. Radio and television are examples of broadcast mode of communication.

Analog Communication

Analog communication is the process of conveying (sending, receiving, and processing) of information including image, voice and video by using analog signals. In analog comunication, the data is transferred with the help of analog signal in between transmitter and receiver as shown in Figure 12.17. Analog communication uses an analog signal which varies in amplitude, phase, or some other property with time. For example, a sinusoidal signal is continuous in nature. It has continuous amplitude with continuous time.

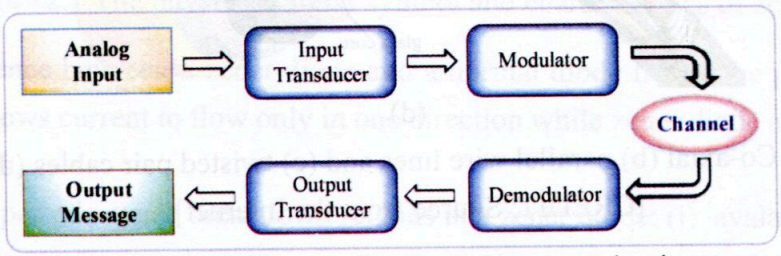


Figure 12.17 Block diagram for analog communication system

Digital Communication

There are many types of digital communication, commonly referred to as digital communication channels. These include email, phone calls, video conferencing and many types of instant messaging like short message service (SMS) and web chats. Transfer of data occurs in the form of

digital bit stream, i.e. 0 or 1 over a point-to-point (or) point-to-multipoint transmission medium. The block diagram of digital communication system is shown in Figure 11.18. Data can be broken into packets as discrete messages which is not allowed in analog communication. With digital communication, transmission errors can be detected and corrected.

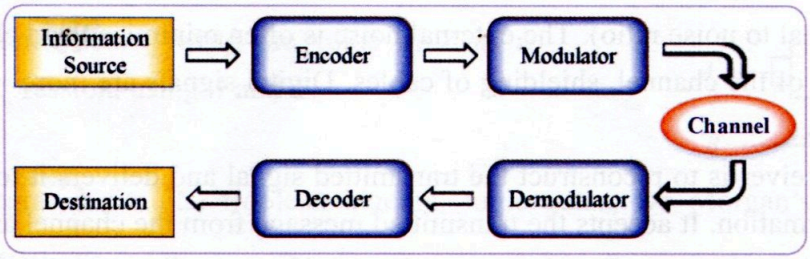


Figure 12.18 Block diagram for digital communication system

Wired (Line) Communication

In line communication, there is a physical connection between source and destination. The wired connections between two points are known as transmission lines. The wires that are most popular for wired communication are: co-axial, parallel wire lines, twisted pair cables and optical fibre as shown in Figure 12.19.

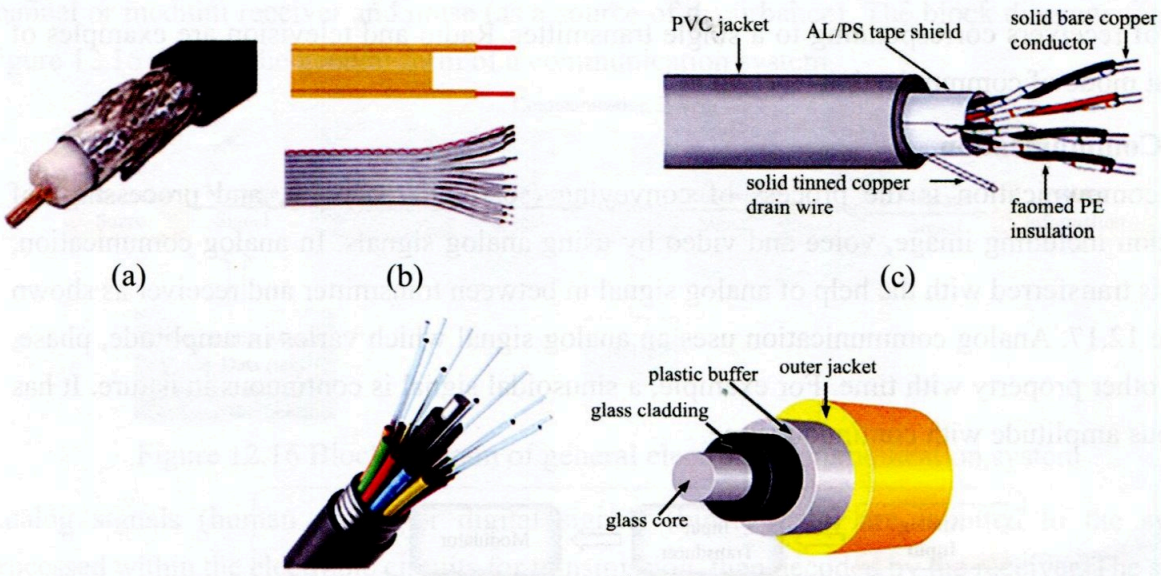


Figure 12.19 (a) Co-axial (b) parallel wire lines and (c) twisted pair cables (d) optical fibre [CREDIT: Source from the Internet]

Wireless (Space) Communication

The simplest space communications rely on two things: a transmitter and a receiver. A transmitter encodes a message onto electromagnetic waves through modulation, which changes properties of the wave to represent the data. These waves flow through space toward the receiver. Figure 12.20 shows the space communication system.

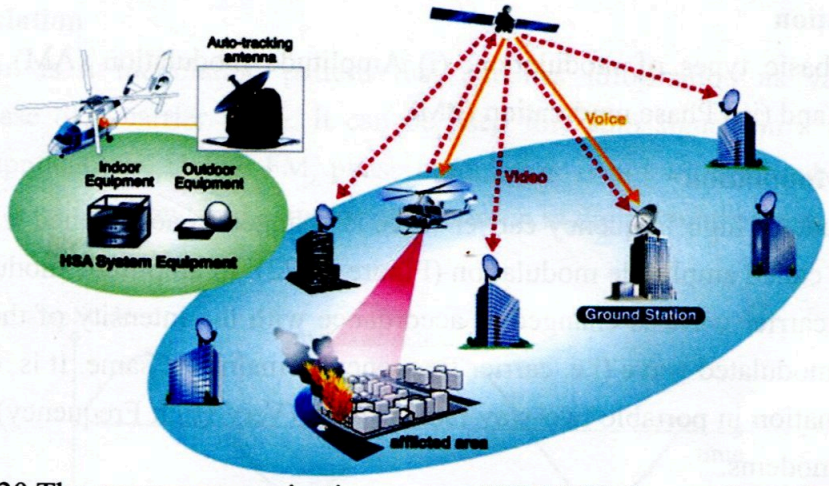


Figure 12.20 The space communication system [CREDIT: Source from the Internet]

Transducer

A transducer is a device for converting a non-electrical input into an electrical signal (or) vice versa. The process of converting energy from one form to another is known as transduction. Transducer contains two parts that are closely related to each other, i.e. the sensing element and transduction element. The sensing element is called the sensor. Sensor is a device which produces a measurable response due to a change in physical conditions. The transduction element converts the sensor output to suitable electrical form as shown in Figure 12.21.

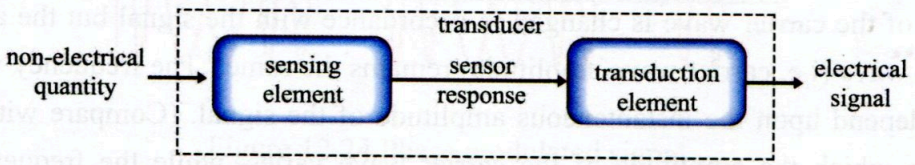


Figure 12.21 Block diagram of a transducer

In general, a transducer is a device that transforms a signal from one energy form to another energy form. Some common examples of transducers include loudspeakers, microphones, oximeter, thermometers, LEDs and antenna.

Modulation

The process of changing some characteristics (e.g. amplitude, frequency or phase) of a carrier wave (e.g. radio wave) in accordance with the intensity of the signal is known as modulation. The resultant wave is called modulated wave, and contains the audio signal. In modulation, the sound wave is electrically impressed on an electromagnetic wave.

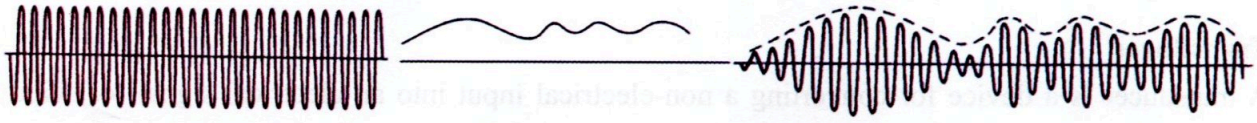
Modulation permits the transmission to occur at high frequency while it simultaneously allows the carrying of the audio signal. It is also the process of manipulating the amplitude, the frequency or phase of a carrier wave in response to an incoming voice, video or data signal.

Types of Modulation

There are three basic types of modulation: (i) Amplitude modulation (AM), (ii) Frequency modulation (FM) and (iii) Phase modulation (PM).

(i) **Amplitude Modulation**

When the amplitude of high frequency carrier wave is changed in accordance with the intensity of the signal, it is called amplitude modulation (Figure 12.22). In amplitude modulation, only the amplitude of the carrier wave is changed in accordance with the intensity of the signal but the frequency of the modulated wave (i.e. carrier frequency) remains the same. It is commonly used to transmit information in portable two-way radios, VHF (Very High Frequency), aircraft radios and in computer modems.



(a) (b) (c)

Figure 12.22 (a) unmodulated carrier wave (b) sound wave (c) amplitude modulated wave (AM)

(ii) **Frequency Modulation**

When the frequency of carrier wave is changed in accordance with the intensity of the signal, it is called frequency modulation (FM) as shown in Figure 12.23. In frequency modulation, only the frequency of the carrier wave is changed in accordance with the signal but the amplitude of the modulated wave (i.e. carrier wave amplitude) remains the same. The frequency variations of carrier wave depend upon the instantaneous amplitude of the signal. (Compare with amplitude modulation, in which the amplitude of the carrier wave varies, while the frequency remains constant.) It is used in radio, telemetry, radar and seismology. FM is widely used for broadcasting music and speech, two-way radio systems, magnetic tape-recording systems and some video-transmission systems.

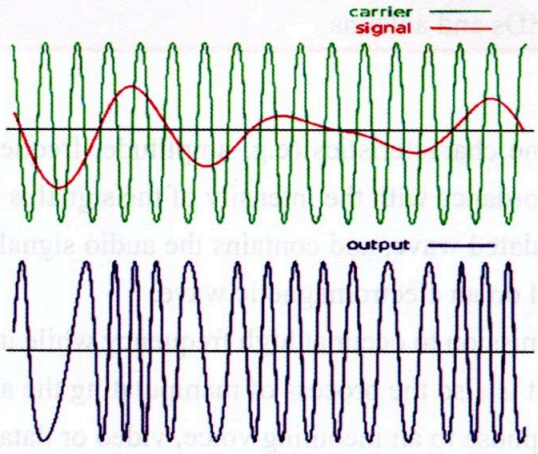


Figure 12.23 Frequency modulated signal

(iii) **Phase Modulation**

Phase modulation is a modulation pattern that encodes information as variations in the instantaneous phase of a carrier wave. It can be used for radio signals in a variety of radio communication applications. Unlike FM, phase modulation is not widely used for transmitting radio waves. It is used for signal and waveform generation in digital synthesizers (Figure 12.24).

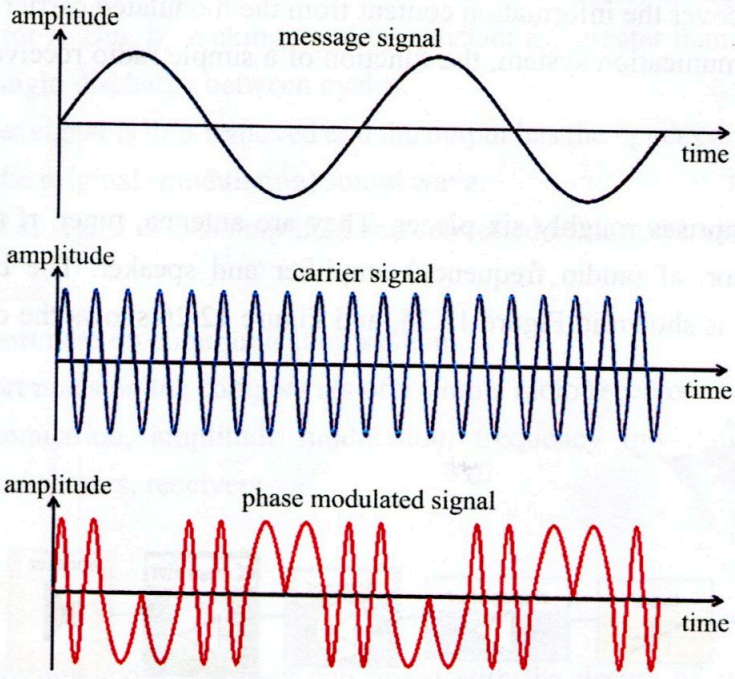


Figure 12.24 Phase modulated signal

Table 12.3 The comparison of frequency modulation (FM) and amplitude modulation (AM)

FM	AM
(i) The amplitude of carrier wave remains constant with modulation.	(i) The amplitude of carrier wave changes with modulation.
(ii) The carrier frequency changes according to the strength of the modulating signal.	(ii) The carrier amplitude changes according to the strength of the modulating signal.
(iii) The carrier frequency changes with modulation.	(iii) The carrier frequency remains constant with modulation.
(iv) FM radio ranges in a higher spectrum from 88 to 108 MHz. (or) 1200 to 2400 bits per second.	(iv) AM radio ranges from 535 to 1705 kHz (or) up to 1200 bits per second.

Demodulation

The process of recovering the audio signal from the modulated wave is known as demodulation or detection. At the broadcasting station, modulation is done to transmit the audio signal over longer distances to a receiver. When the modulated wave is picked up by the radio receiver, it is necessary to recover the audio signal from it. This process is accomplished in the radio receiver and is called demodulation. A demodulator is an electronic circuit (or computer program in a software defined radio) that is used to recover the information content from the modulated carrier wave.

As an example of communication system, the function of a simple radio receiver is described as follows.

Simple Radio Receiver

The radio receiver comprises roughly six pieces. They are antenna, tuner, rf (radio frequency) amplifier, diode detector, af (audio frequency) amplifier and speaker. The bolck diagram of a simple radio receiver is shown in Figure 12.25, and Figure 12.26 shows the circuit diagram of AM radio receiver.

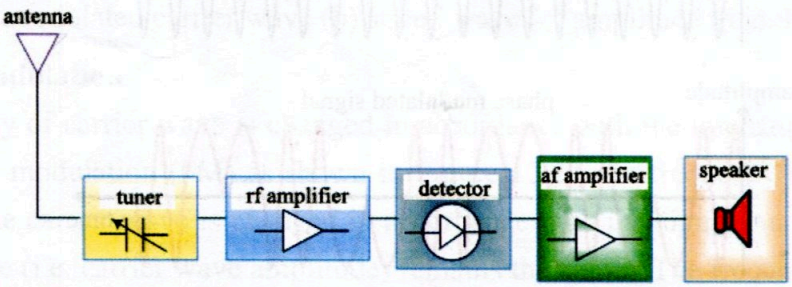
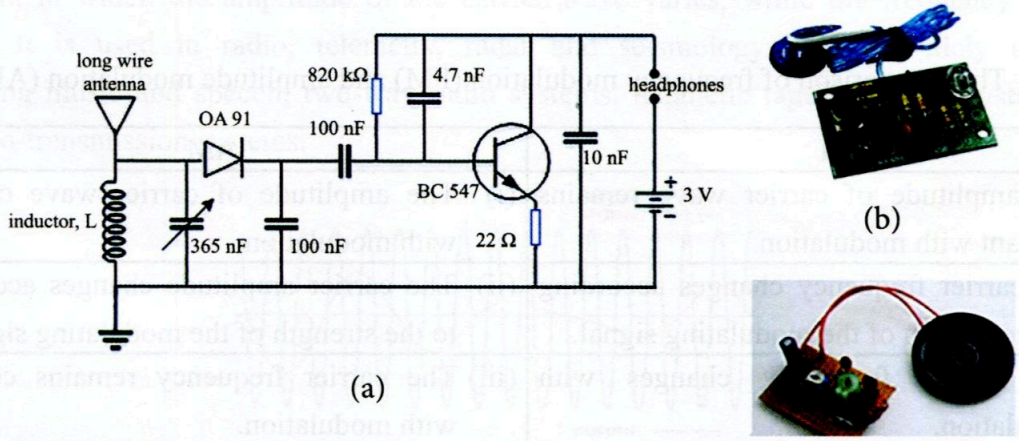


Figure 12.25 Block diagram of a simple radio receivers



(a)

(b)

(c)

Figure 12.26 (a) The circuit diagram of AM radio receiver
(b) Printed board of AM radio receiver
(c) Printed board of FM radio receiver

The receiving antenna picks up the radio signals which is coupled to tuning circuit (through the inductor L) and by adjusting a variable capacitor, a particular radio frequency can be selected. The weak signal is amplified and fed into the diode detector.

The input signal contains an *rf* (radio frequency) component, the carrier, and an *af* (audio frequency) component, the sound signal.

During each carrier cycle, the diode turns on briefly and charges the capacitor in the detection circuit to the peak voltage of the particular cycle. Between cycles, the capacitor would discharge through the resistor R. But, by making the time constant RC greater than the period of the carrier wave, there is a slight discharge between cycles.

Most of the carrier signal is thus removed and the output has the upper part of the modulated wave which is that of the original (modulating) sound wave.

The demodulated af signal is then amplified and converted back into sound by a speaker.

Reviewed Exercise

1. Give a short note on communication system.
2. Write short notes on the components of a simple radio receiver.

Key Words: modulation, amplitude modulation, frequency modulation, phase modulation, transmitters, receivers

SUMMARY

The branch of physics and technology concerned with the design of circuits using transistors, integrated circuits and microchips, and also with the behaviour and movement of electrons in a semiconductor, conductor or vacuum is known as **electronics**.

Digital electronics is a field of electronics involving the study of digital signals and the engineering of devices that use or produce them. This is in contrast to analog electronics and analog signals.

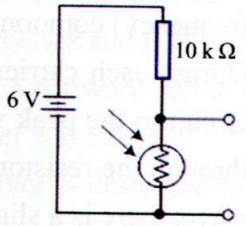
The process of changing some characteristics (e.g. amplitude, frequency or phase) of a carrier wave (e.g. radio wave) in accordance with the intensity of the signal is known as **modulation**.

The process of recovering the audio signal from the modulated wave is known as **demodulation** or detection.

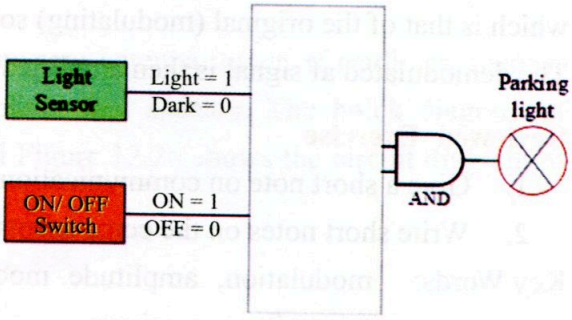
EXERCISES

1. Which electrical component is used to control the brightness of street lamps?
   (i) LED (ii) LDR (iii) thermistor

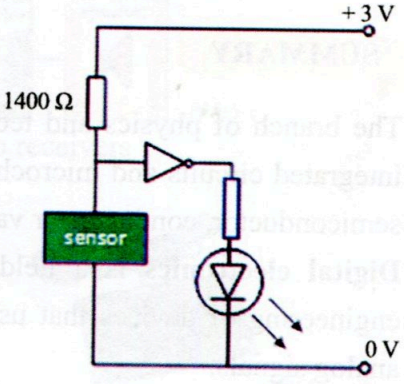
2. A potential divider consists of an LDR and a resistor connected to a 6 V battery in following figure. What should be the resistance of the LDR for the output to be 2.5 V?



3. A car may be fitted with an automatic parking light which switches ON when the car is parked at night. Figure shows an incomplete system for a parking light. Copy and complete the figure to show how a NOT gate can be used to make the parking light work. Use the correct symbol for a NOT gate.



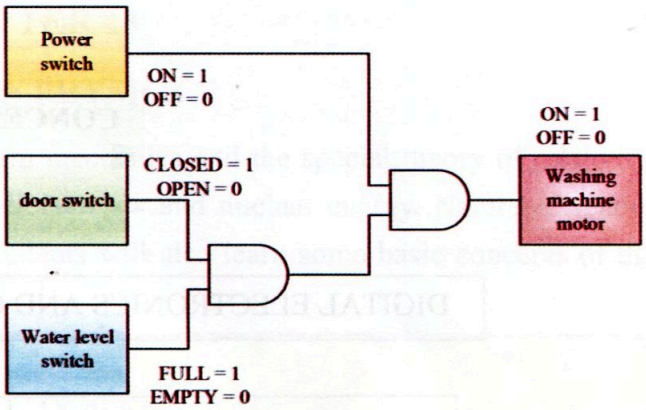
4. A student uses the sensor in the circuit shown in figure. The LED lights up when the temperature reaches 40 °C. Name a suitable temperature sensor and draw its circuit symbol.



5. In the giving table, *P*, *Q*, *R* and *S* are the outputs of four logic gates. Name the logic gate which represents each output.

input		output			
		P	*Q*	*R*	*S*
0	0	0	0	1	1
0	1	0	1	1	0
1	0	0	1	1	0
1	1	1	1	0	0

6. Figure shows a control system which may be fitted in an automatic washing machine. What conditions will (i) start (ii) stop the washing machine working?



7. Simplify the following expressions if possible:
   (i) $A + AB + A\bar{B}C$ (ii) $(\bar{A} + B)C + ABC$
8. Draw the logic circuit represented by each of the following expressions:
   (i) $A + B + C$ (ii) ABC (iii) $AB + C$ (iv) $AB + CD$
9. Construct a truth table for each of the followings:
   (i) $AB + BC$ (ii) $(A + B)C$ (iii) $(A + B)(\bar{B}+C)$
10. Find the value of X for all possible values of variables.
    (i) $X = (A + B)C + B$ (ii) $\boldsymbol{X = \overline{(A + B)}C}$ (iii) $X = A\bar{B}C + AB$
11. Apply De Morgan's theorems to the expression $\boldsymbol{\overline{ABC}}$ and $\boldsymbol{\bar{A} + \bar{B} + \bar{C} + \bar{D}}$.
12. Can sound travel very far? Why? Which device can be used to transmit sounds over long distance? What range of frequencies do the radio waves have?
13. What is meant by "AM and FM"?
14. What are the differences between FM and AM? Explain with illustrative diagrams?
15. Which range of frequency a radio telephone or walkie talkie is communicating? Also find the corresponding wavelength.

CONCEPT MAP

DIGITAL ELECTRONICS AND COMMUNICATION SYSTEMS

Electronic System

- zener diode
- light emitting diode (LED)
- photodiode
- light dependent resistor (LDR)
- thermistor
- liquid crystal display (LCD)

Digital Electronics

Combination of logic gates
- fire alarm system
- street lighting system

Boolean algebra
- basic laws
- basic rules

De Morgan's theorem

Electronic Communication System

Modes of Communication System
- analog communication
- digital communication
- wired (Line) communication
- wireless communication

Modulation
- amplitude modulation (AM)
- frequency modulation (FM)
- phase modulation (PM)

Demodulation

Simple radio receiver

CHAPTER 13

MODERN PHYSICS

This chapter describes nuclear physics, quantum mechanics and the special theory of relativity. Nuclear physics includes the radioactivity, radioisotopes and nuclear energy. Nature of matter wave will be learnt in quantum mechanics. Students will also learn some basic concepts of the special theory of relativity.

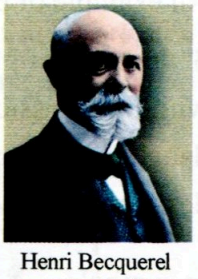
Learning Outcomes

It is expected that students will

- understand radioactivity, radioactive decays, alpha, beta, gamma rays and their properties.
- investigate activity and half-life of a radioactive substance.
- realize some application of radioisotopes.
- examine fission, nuclear energy, its peaceful use as a nuclear reactor and the environmental impacts.
- realize the de Broglie matter wave associated with wave particle dual nature and the quantization of angular momentum in theory of hydrogen atom.
- describe the experiments and scientific problems that led Albert Einstein to develop the special theory of relativity.
- understand the postulates on the special theory of relativity.
- describe the relativistic effects: time dilation, length contraction, and relativistic momentum.
- explain and perform calculations involving mass-energy equivalence.

13.1 RADIOACTIVITY AND USES OF RADIOACTIVITY

In 1896, Henry Becquerel discovered that uranium salts emit radiations which affect photographic plates and cause ionization. The radiations are emitted from the nucleus of uranium atoms and are of three types; namely alpha, beta and gamma rays. This phenomenon is known as radioactivity.



Henri Becquerel
(1852-1908)

Hence, the emission of some or all of alpha, beta and gamma rays from the nucleus of an unstable atom is called radioactivity. The term radioactivity is also known as radioactive decay, radioactive disintegration, nuclear decay, or nuclear disintegration.

Of all 92 elements which exist in nature, only a few heavy elements with atomic number greater than 83 are radioactive or unstable. Some examples of radioactive elements (or radioactive substances) are uranium, thorium, radium, polonium and radon; of which radium and polonium were discovered by Marie Curie. Radioactive elements can also be artificially produced using nuclear reactors and particle accelerators.

Marie Curie
(1867-1934)

1. Alpha Decay, Beta Decay and Gamma Decay

(i) Alpha decay

Alpha decay (α decay) is a process in which an atomic nucleus emits an alpha particle (helium nucleus) and transforms (or decays) into a different atomic nucleus, with mass number reduces by four and atomic number reduces by two as described by Eq. (13.1).

In symbols, $${}^{A}_{Z}X \rightarrow {}^{A-4}_{Z-2}Y + {}^{4}_{2}He \qquad (13.1)$$

where ${}^{A}_{Z}X$ = parent nucleus, ${}^{A-4}_{Z-2}Y$ = daughter nucleus, and ${}^{4}_{2}He$ = alpha particle.

An example of alpha decay is $${}^{238}_{92}U \rightarrow {}^{234}_{90}Th + {}^{4}_{2}He$$

Note that atomic number and mass number are conserved in the decay process.

(ii) Beta decay

There are two types of beta decay: namely, β^- decay and β^+ decay.

β^- decay (beta-minus decay) is a process in which a neutron in an atomic nucleus changes to a proton with the emission of an electron (β^-) and an antineutrino ($\bar{\nu}_e$). The parent nucleus transforms into the daughter nucleus, with mass number remains unchanged but atomic number increases by one as described by Eq.(13.2).

In symbols, $${}^{A}_{Z}X \rightarrow {}^{A}_{Z+1}Y + {}^{0}_{-1}e + \bar{\nu}_e \qquad (13.2)$$

where ${}^{A}_{Z}X$ = parent nucleus, ${}^{A}_{Z+1}Y$ = daughter nucleus,

${}^{0}_{-1}e$ = electron (β^- particle), $\bar{\nu}_e$ = antineutrino

An example of β^- decay is $${}^{14}_{6}C \rightarrow {}^{14}_{7}N + {}^{0}_{-1}e + \bar{\nu}_e$$

β^+ decay (beta-plus decay) is a process in which a proton in an atomic nucleus changes to a neutron with the emission of a positron (β^+) and a neutrino (ν_e). The parent nucleus transforms into the daughter nucleus, with mass number remains unchanged but atomic number decreases by one as described by Eq. (13.3).

In symbols, $${}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + {}^{0}_{+1}e + \nu_e \qquad (13.3)$$

where ${}^{A}_{Z}X$ = parent nucleus, ${}^{A}_{Z-1}Y$ = daughter nucleus,

${}^{0}_{+1}e$ = positron (β^+ particle), ν_e = neutrino

An example of β^+ decay is $${}^{11}_{6}C \rightarrow {}^{11}_{5}B + {}^{0}_{+1}e + \nu_e$$

Positron, the antiparticle of electron, has the same mass and same magnitude of electric charge as an electron; but positron is positively charged. The existence of positron was theoretically predicted by Dirac in 1931 and experimentally detected by Anderson in 1932.

Neutrino, a fundamental particle as an electron, has no charge and negligibly small rest mass ($< 2 \times 10^{-37}$ kg). There are three types of neutrino; called electron neutrino (ν_e), the muon neutrino (ν_μ) and the tau neutrino (ν_τ). Each neutrino is associated with an antineutrino: $\bar{\nu}_e$, $\bar{\nu}_\mu$ and $\bar{\nu}_\tau$.

For a nucleus to be stable, there must be an appropriate ratio of protons and neutrons. If a nucleus has excess neutrons (e.g. $^{14}_{6}C$) β^- decay takes place. Conversely, a nucleus with excess protons (e.g. $^{11}_{6}C$) undergoes β^+ decay.

(iii) **Gamma decay**

In gamma decay (γ decay), the daughter nucleus resulting from alpha or beta decay is left in an excited state and it can then decay to a lower energy state by emitting a gamma ray photon. In gamma decay, the atomic number and mass number of the parent and daughter nuclei remain unchanged; only the parent nucleus releases energy in the form of electromagnetic radiation as described by Eq. (13.4).

In symbols,
$$^{A}_{Z}X^{*} \rightarrow {}^{A}_{Z}X + {}^{0}_{0}\gamma \qquad (13.4)$$

where $^{A}_{Z}X^{*}$ = parent nucleus (excited state),

$^{A}_{Z}X$ = daughter nucleus (ground state), $^{0}_{0}\gamma$ = gamma photon

For example, cobalt-60 undergoes β^- emission to an excited state of nickle-60, which in turn decays to the ground state by γ emission.

$$^{60}_{27}Co \rightarrow {}^{60}_{28}Ni^{*} + {}^{0}_{-1}e + \bar{\nu}_e , \qquad {}^{60}_{28}Ni^{*} \rightarrow {}^{60}_{28}Ni + {}^{0}_{0}\gamma$$

Hence, gamma decay takes place when a nucleus at an excited state deexcites to a lower excited state or to the ground state.

The photon is the quantum of the electromagnetic field including electromagnetic radiation such as light, X-rays and gamma rays. Photons are massless and move at the speed of light in vacuum. The energy of a photon is inversely proportional to the wavelength of the electromagnetic radiation associated with it.

α decay, β^- decay, β^+ decay and γ decay are illustrated in the following Figure 13.1.

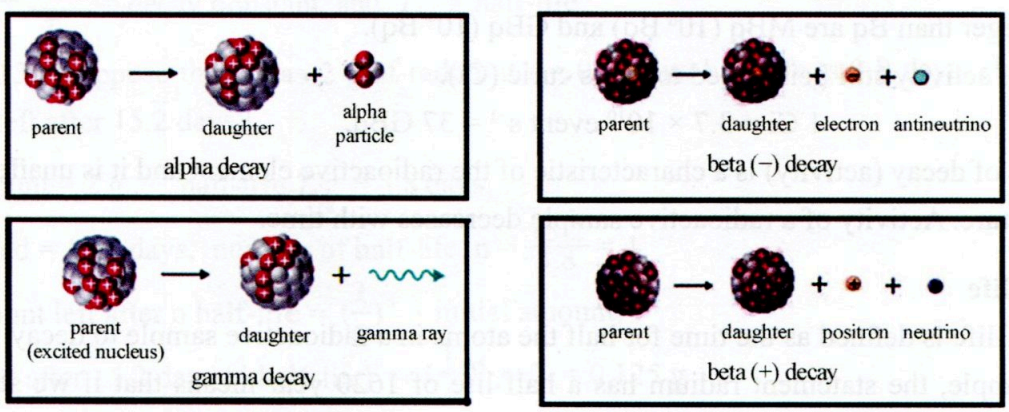


Figure 13.1 Illustrations of α decay, β^- decay, β^+ decay and γ decay

2. **Properties of alpha rays, beta rays and gamma rays**

(i) **Properties of alpha rays**

- Alpha rays consist of helium nuclei, positively charged particles. An α particle consists of two protons and two neutrons. The charge of an α particle is $+2e$ ($+3.2 \times 10^{-19}$ C).
- Alpha rays can be deflected by electric and magnetic fields.
- Alpha rays are the most ionizing of the three rays.
- Alpha rays are the least penetrating of the three rays.
- Alpha rays can be stopped by a thick sheet of paper.

(ii) **Properties of beta rays**

- Beta rays consists of either electrons (charge $= -e$) or positrons (charge $= +e$) moving with varying speeds.
- Beta rays can be deflected by electric and magnetic fields.
- Beta rays are less ionizing than alpha rays.
- Beta rays are more penetrating than alpha rays.
- Beta rays can be stopped by a few millimetres of aluminium.

(iii) **Properties of gamma rays**

- Gamma rays are electromagnetic waves like light and X-rays, but have much shorter wavelengths. Gamma rays are high energy photons. Gamma rays have no charge.
- Gamma rays cannot be deflected by electric and magnetic fields.
- Gamma rays are the least ionizing of the three rays.
- Gamma rays are the most penetrating of the three rays.
- Intensity of gamma rays can be greatly reduced by several centimetres of lead but they are never completely stopped.

3. **Activity**

The rate of decay of a radioactive sample is called activity. It is the rate at which alpha, beta particles or gamma rays are emitted.

SI unit for activity is becquerel (Bq). 1 Bq = 1 event s^{-1} (or 1 disintegration s^{-1})

Units larger than Bq are MBq (10^6 Bq) and GBq (10^9 Bq).

A unit of activity still being used today is curie (Ci).

$$1 \text{ Ci} = 3.7 \times 10^{10} \text{ event s}^{-1} = 37 \text{ GBq}.$$

The rate of decay (activity) is a characteristic of the radioactive element and it is unaffected by the temperature. Activity of a radioactive sample decreases with time.

4. **Half-life**

The half-life is defined as the time for half the atoms in a radioactive sample to decay.

For example, the statement radium has a half-life of 1620 year means that if we start with N_0

atoms of radium, then only $\frac{N_0}{2}$ atoms will remain after a time of 1620 years has elapsed. (After another 1620 years only $\frac{N_0}{4}$ atoms will remain and so on.)

Half-life of radioactive substances ranges from about femtosecond (10^{-15} s) to billions of year.

Graphical illustration of half-life

Consider radon of half-life $T_{½}$ is 3.8 days.

time (days)	mass of radon (g)
0.0	1
3.8	$\frac{1}{2}$
7.6	$\frac{1}{4}$
11.4	$\frac{1}{8}$

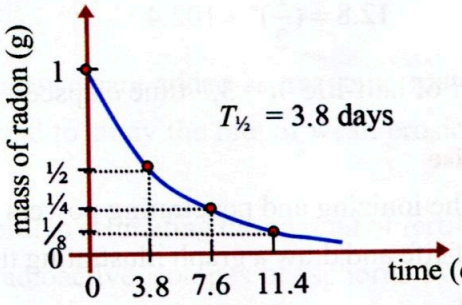


Figure 13.2 A graphical illustration of half-life or exponential nature of radioactive decay

$$\text{Amount left after n half-life} = (\frac{1}{2})^n \times \text{initial amount.} \qquad (13.5)$$

Radioactive decay is a statistical process: it is impossible to predict when a particular atom will decay, regardless of how long the atom has existed. However, for a significant number of identical atoms, the decay rate can be expressed as a decay constant λ or as half-life $T_{1/2}$. This leads to the exponential law of radioactive decay given by the relation as below.

$$N_t = N_0\, e^{(-\lambda t)} = N_0\, e^{(-\frac{0.693}{T_{½}} t)} \qquad (13.6)$$

where N_t = number of radioactive atoms at time t,

N_0 = initial number of radioactive atoms,

$\lambda = \frac{0.693}{T_{½}}$ = decay constant, and $T_{1/2}$ = half-life

Example 13.1 Suppose there was 2 g of radon at $t = 0$ and its half-life is 3.8 days. How much would be left after 15.2 days?

Initial amount = 2 g , half-life $T_{½}$ = 3.8 days,

time elapsed = 15.2 days, number of half-life $n = \frac{15.2}{3.8} = 4$

Since amount left after n half-life = $(\frac{1}{2})^n \times$ initial amount,

amount left after 15.2 days (4 half-life) = $(\frac{1}{2})^4 \times 2 = 0.125$ g

Example 13.2 A radioisotope has a half-life of 20 min. If there is initially 102.4 g of this isotope, how much time elapses for 12.8 g to be left?

Initial amount =102.4 g, amount left =12.8 g, half-life $T_{1/2}$ = 20 min,

Since, $$\frac{12.8}{102.4} = \frac{1}{8} = (\frac{1}{2})^3$$

$$\text{amount left} = (\frac{1}{2})^n \times \text{initial amount}$$

$$12.8 = (\frac{1}{2})^n \times 102.4$$

Therefore, number of half-life $n = 3$, time elapsed = $n \times T_{1/2} = 3 \times 20 = 60$ min = 1 h

Reviewed Exercise

1. Compare the ionizing and penetrating powers of alpha, beta and gamma rays.
2. Define half-life and draw a graph illustrating the half-life of radium which has a half-life of 1620 years.
3. Cobalt-60 (half-life = 5.25 years) is often used as a radiation source in medicine. How long after a new sample is delivered, will the activity have decreased to $\frac{1}{8}$ of its original value?
4. Suppose that initially there was 100 μCi of iodine-131 and its half-life is 8 days. How much would it be left after 24 days?

Key Words: radioactivity, α decay, β decay, γ decay, activity, half-life

5. Uses of radioactivity (or) uses of radioisotopes

A radioisotope (radioactive isotope or radionuclide) is an atom with an unstable nucleus which undergoes radioactive decay, resulting in the emission of alpha, beta and gamma rays. Radioisotopes with suitable half-lives play an important part in a number of technologies. Various uses of radioisotopes are described as (i) tracer (ii) radiotherapy (iii) industrial gamma radiography (iv) thickness monitor (v) radioactive dating and (vi) gamma irradiation of seeds and foods.

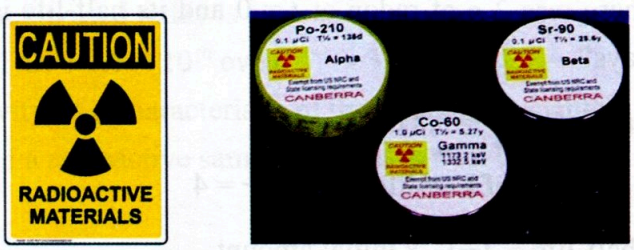


Figure 13.3 Radioactive material sign and some radioactive sources

(i) **Tracer**

A tracer is a small amount of radioactive isotope introduced into a system in order to follow the behaviour of some components of that system.

In medicine, radioactive tracers are effectively used in diagnostic procedures. For example, to check thyroid function, the patient drinks an iodine solution containing a small amount of radioactive iodine-131(gamma emitter). Over the next 24 hours, a gamma ray detector measures the activity of the iodine tracer to find out how quickly it becomes concentrated in the thyroid gland. For tests like those above, artificial radioisotopes with short half-lives are used so that there is no detectable radiation after a few days.

In industry, small amounts of radioactive substances are added as tracers to materials in various processes to monitor fluid flow, detect leaks, and to study the rate of wear, erosion and corrosion of equipment.

In agriculture, radioisotope tracers are very useful in estimating the amount of fertilizers that should be supplied to soil. Fertilizers labelled with radioactive isotopes phosphorus-32 and nitrogen-15 have been used to study the uptake, retention and utilization of phosphate and nitrogen fertilizers.

(ii) **Radiotherapy**

Treatment with radiation is known as radiotherapy. Radioisotopes, such as cobalt-60 and cesium-137, are widely used to treat cancer in radiotherapy. Gamma rays from cobalt-60 can penetrate deep into the body and kill living cells. A highly concentrated beam from a cobalt-60 source can be used to kill cancer cells in a tumour as shown in Figure 13.4.

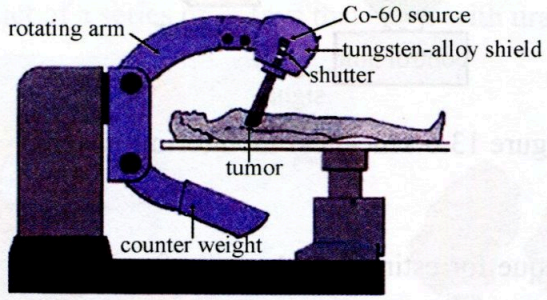


Figure 13. 4 Cobalt-60 radiotherapy

(iii) **Industrial gamma radiography**

In industrial gamma radiography, radioactive materials are used to inspect metal castings or welded joints. Radioactive gamma emitter in a sealed capsule is placed on one side of the object being screened, and photographic film is placed on the other side as shown in Figure 13.5.

The gamma rays pass through the object and create an image on the film which shows flaws in the object. Radioisotopes have the advantage that they can be taken to the site when an examination is required and no power is needed. However, they must be properly shielded when in use and also at other times.

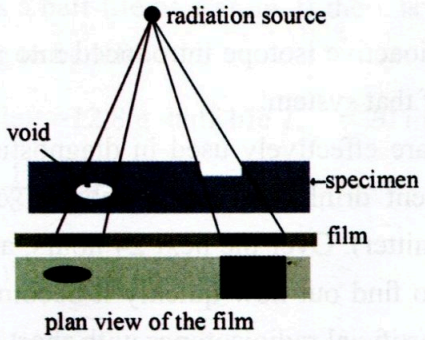


Figure 13.5 Industrial gamma radiography

(iv) **Thickness monitor**

The ability to use radioisotopes to accurately measure thickness is widely used as thickness monitoring system in the production of sheet materials, including metal, textiles, paper, plastics, and others.

In a thickness monitoring system shown in Figure 13.6, the moving band of tyre cord has a beta source on one side and a detector on the other. If, for example, the cord from the rollers becomes too thin, more beta radiation reaches the detector. Detector sends signal to the control unit which adjust the gap between the rollers.

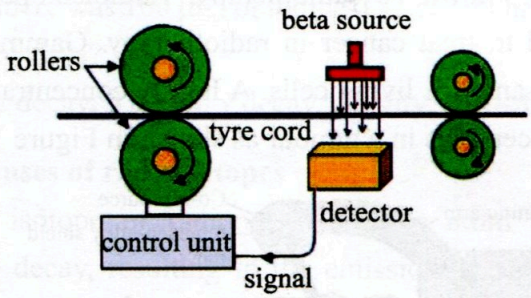


Figure 13.6 Thickness monitoring system

(v) **Radioactive dating**

Radioactive dating is a technique for estimating the age of an object by measuring the amounts of various radioisotopes present in it. Natural radionuclides are used in archaeology, geology, and paleontology to measure ages of fossil materials, rocks and minerals.

Carbon dating: Carbon dating is a technique, developed by Willard Libby, for estimating the age of the remains of plants and animals by measuring the activity of carbon-14 presents in the remains. Carbon is present in atmosphere and also in the bodies of plants and animals. A small proportion is radioactive carbon-14 (half-life 5730 years).

Although carbon-14 undergoes radioactive decay, it is also continually produced by cosmic rays entering the upper atmosphere. High energy neutrons of cosmic rays bombard nitrogen-14 atoms and produce carbon -14 atoms. Thus, the abundance of carbon-14 in the atmosphere is essentially constant. There is one carbon-14 atom in every trillion (10^{12}) of carbon-12 atoms.

When the plants and animals are alive, they absorb and give out carbon. There is equilibrium of carbon between the atmosphere and the bodies of plant and animals as shown in Figure 13.7. Thus, the amount of carbon-14 present in the living body is essentially constant.

However, when an animal or a plant dies carbon cycle stops and the carbon-14 in the remain gradually decreases by radioactive decay. By measuring the activity of carbon-14, the age of the remain can be estimated.

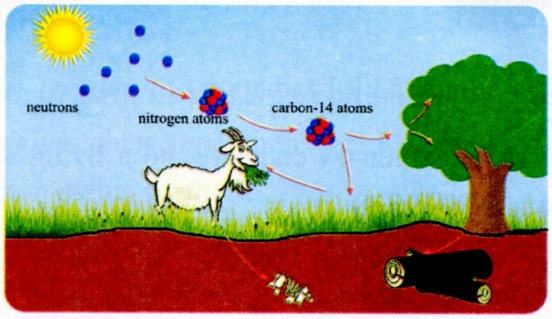


Figure 13.7 Accumulation of carbon-14 in plants and animals

Dating of rocks: The age of the rocks can also be estimated by radioactive dating technique. When rocks are formed, some radioisotopes become trapped in them. For example, potassium-40 (radioactive) is trapped in igneous rocks. As potassium-40 decays, more and more of its stable decay product, argon-40 is created as shown in Figure 13.8. Provided that argon has not escaped from the rock, the age of the rock can be estimated from the proportion of potassiun-40 to argon-40. Igneous rocks can also be dated by the proportion of uranium to lead isotopes. Lead is the stable final decay product of a series of decays that starts with uranium.

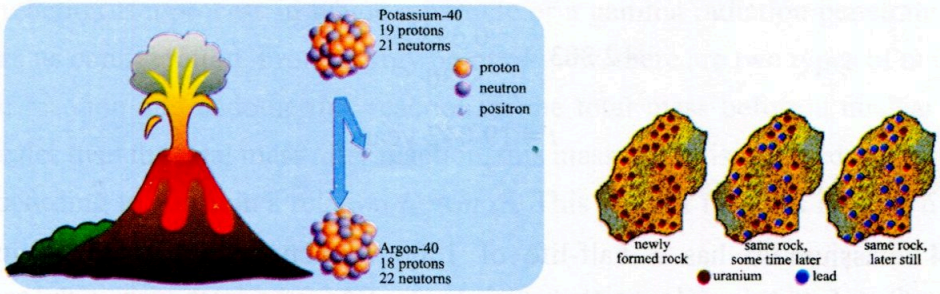


Figure 13.8 Radioactive dating of rocks

(vi) **Gamma irradiation of seeds and foods**

Gamma and neutron irradiation of seeds can be used to induce mutations to develop desired mutants that are resistant to disease, produce a higher yield and allow earlier ripening.

Radiation can also be used to destroy microbes in food and control insect and parasite infestation in harvested food. Irradiation of food has potential to produce safe foods with long shelf life. Certain seeds and canned food can be stored for longer periods by gently exposing them to radiations as shown in Figure 13.9.

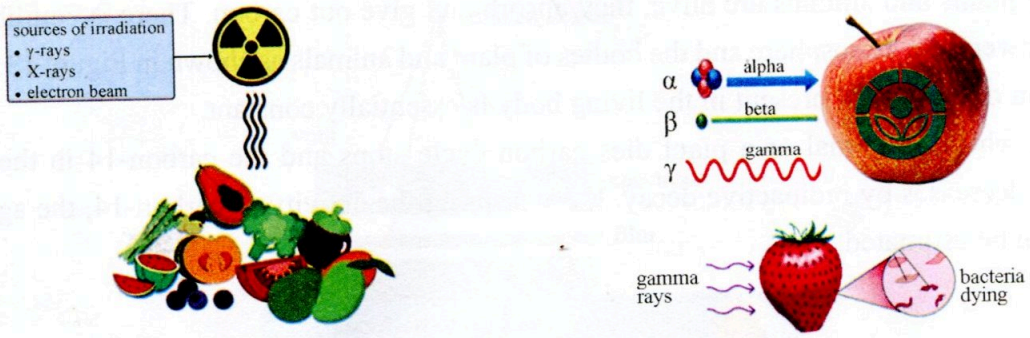


Figure 13.9 Irradiation of foods

Example 13.3 If a fossil has 10 % carbon-14 compared to a living sample, find the age of the fossil. (half-life of carbon-14 = 5730 years)

$\frac{N_t}{N_0} = 0.1$, $T_{1/2} = 5730$ years

We have the relation

$$N_t = N_0 \, e^{(-\frac{0.693}{T_{1/2}} t)}$$

$$\ln(\frac{N_t}{N_0}) = -\frac{0.693}{T_{1/2}} t$$

$$\ln 0.1 = -\frac{0.693}{T_{1/2}} t$$

$$2.303 \log_{10} 0.1 = -\frac{0.693}{T_{1/2}} t$$

$$-2.303 = -\frac{0.693}{5730} t$$

$$t = 19\,042 \text{ years}$$

Age of the fossil is 19 042 years.

Example 13.4 Potassium-40 has a half-life of 1300 million years. If the atomic ratio of potassium-40 to argon-40 in a rock sample is found to be 1:3, estimate the age of the rock. Assume that there was initially no argon in the rock and argon formed by the decay of potassium is trapped in the rock.

$T_{1/2}$ = 1300 million years, the atomic ratio of potassium-40 to argon-40 in rock sample = $\frac{1}{3}$

Assume that there were N_0 atoms of potassium-40 when the rock was formed.

After n half-life, the number of potassium-40 atoms left = $(\frac{1}{2})^n N_0$

The number of argon-40 atoms formed = $N_0 - (\frac{1}{2})^n N_0 = N_0 [1 - (\frac{1}{2})^n]$

The atomic ratio of potassium-40 to argon-40 $= \dfrac{(\frac{1}{2})^n}{1-(\frac{1}{2})^n} = \dfrac{1}{2^n-1}$

$$\frac{1}{3} = \frac{1}{2^n-1}$$

Thus , $2^n - 1 = 3$ and $n = 2$, the age of the rock $= 2 \times 1300 = 2600$ million years.

Reviewed Exercise

1. What is carbon dating? How does the proportion of the carbon-14 in the remains of dead plants and animals gives clues about their age? Can the age of living plants and animals be determined by carbon dating method?
2. Aluminium is rolled into sheets 1 mm thick. Of three types of radiations; alpha, beta and gamma, which radiation is suitable for checking the uniformity of thickness. Give the reason for your answer.

Key Words: tracer, radiotherapy, gamma radiography, thickness monitor, carbon dating

13.2 NUCLEAR ENERGY AND ITS ENVIRONMENTAL IMPACT

Of the various form of energy, nuclear energy now provides about 10 % of the world's electricity from about 450 nuclear power reactors. Nuclear power plants currently operate in some 30 countries.

1. Nuclear reaction

A nuclear reaction is a process in which a particle or a gamma radiation penetrate into a nucleus and changes its configuration. From energy point of view, there are two types of nuclear reactions: exothermic reaction and endothermic reaction. If the total mass before a nuclear reaction takes place is greater than the total mass after reaction, this mass defect is released as energy in a nuclear reaction according to Einstein's relation $E = m\,c^2$. This type of reaction is known as exothermic reaction.

On the other hand, if the total mass before a nuclear reaction takes place is less than the total mass after reaction, external energy must be provided for the reaction to take place. This type of reaction is known as endothermic reaction.

Nuclear fission reaction

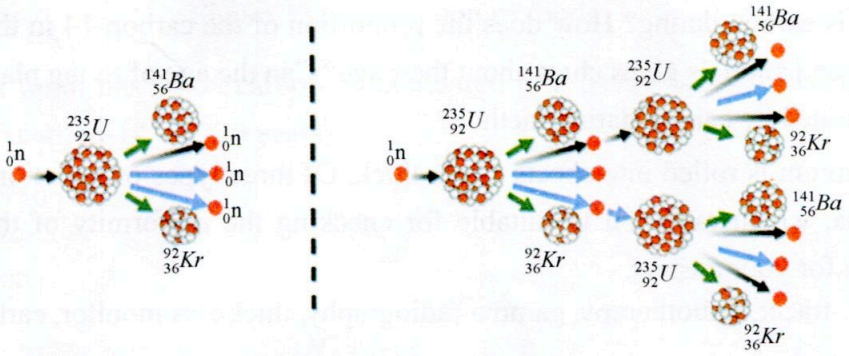
The most important nuclear reaction from which considerable amount of nuclear energy can be extracted is nuclear fission reaction. When a neutron strikes and penetrates a nucleus of uranium-235, the nucleus becomes highly unstable and splits into two lighter nuclei, shooting out two or three neutrons as it does so. This splitting process is called fission, and the fragments are

thrown apart as the energy is released as shown in Figure 13.10 (a).

An example of fission reaction is

$$^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{92}_{36}\text{Kr} + {}^{141}_{56}\text{Ba} + 3\ {}^{1}_{0}\text{n} \qquad (13.7)$$

If the neutrons emitted from a fission reaction go onto split other uranium-235 nuclei and so on, the result is a chain reaction with a huge and rapid release of energy as shown in Figure 13.10 (b). For a chain reaction to be maintained, the uranium-235 has to be above a certain critical mass, otherwise too many neutrons escape and chain reaction cannot occur.



(a) (b)

Figure 13.10 (a) Nuclear fission and (b) fission chain reaction

Energy released from a fission process

We will consider the reaction $^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{92}_{36}\text{Kr} + {}^{141}_{56}\text{Ba} + 3\ {}^{1}_{0}\text{n}$

total mass before reaction = 235.043915 + 1.008665 = 236.052580 u

total mass after reaction = 91.91972 + 140.910575 + 3 × 1.008665

= 235.856290 u

mass defect = 236.052580 − 235.856290 = 0.19629 u

Since, atomic mass unit 1 u = 931.5 MeV

Energy released = 0.19629 × 931.5 = 182.84 MeV

Note that the energy released per atom in a nuclear reaction is about a million times greater than the energy released per atom from a chemical reaction.

2. **Nuclear reactor**

Nuclear reactor is a device in which a fission chain reaction can be initiated, maintained and controlled. In a nuclear power station, the controlled chain reaction takes place in a nuclear reactor and thermal energy (heat) is released at a steady state. This heat energy is used to make steam for turbines which generates electricity, as in a conventional power station.

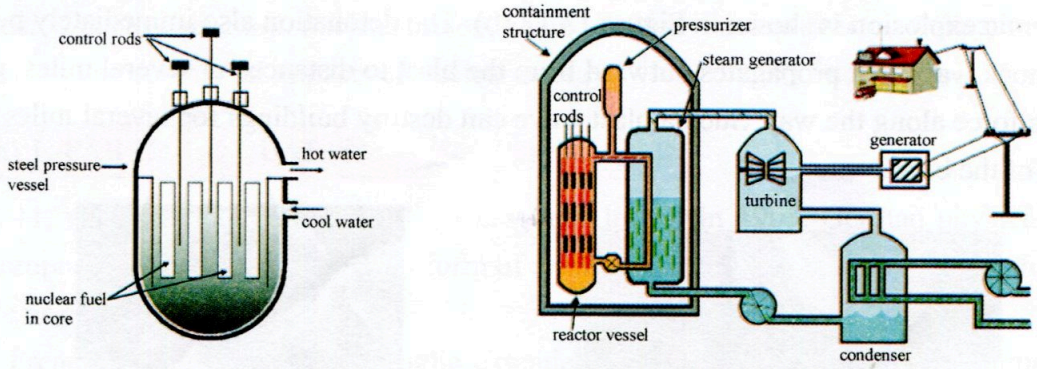
The essential elements of a nuclear reactor are fissionable nuclear fuel, moderator, coolant, control rods, reactor vessel and shielding.

Fissionable nuclear fuel is the material for fission reaction to take place. Uranium-235 or plutonium-239 can be used as fuel.

Neutrons released from a fission reaction are fast neutrons. They need to be slowed down to maintain a chain reaction. Moderator is a material to slow down fast neutrons released from a fission process. Graphite and ordinary water (H_2O) as well as heavy water (D_2O) are used as moderators.

Coolant is to extract the heat produced by fission fragments when they slow down and stop in the reactor material. Water and carbon dioxide CO_2 gas are used as coolant.

Control rods are used to control the rate of reaction by raising or lowering them from the core. Boron and cadmium (neutron absorbing materials) are used as control rods. Reactor vessel is to contain the reactor core, moderator, coolant and control rods. It is a pressurized steel vessel as shown in Figure 13.11(a).



(a) (b)

Figure 13.11 (a) Pressurized water reactor vessel (b) pressurized water reactor system

Shielding (or containment structure) is a protective barrier, usually a dense material. It reduces the passage of radiations from radioactive materials by absorbing them.

A schematic diagram of a pressurized water reactor system is shown in Figure 13.11 (b).

Radioactive waste

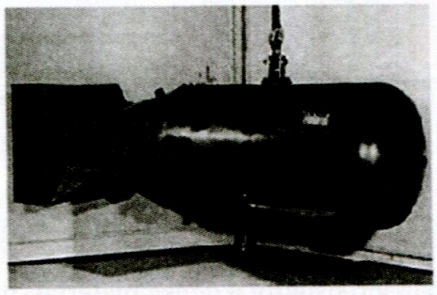
Materials which are radioactive and for which there is no further use are called radioactive waste. After three or four years in a reactor, spent fuel must be removed and replaced with a new one. The amount of uraniun-235 in it has fallen and the fission products has built up. Spent fuel are taken to a reprocessing plant where unused fuel and plutonium are removed. The remaining radioactive waste is sealed off and stored with thick shielding around it. Since some of the isotopes have long half-lives, safe storage will be needed for thousands of years.

Many of the fission products are radioactive and dangerous to be released into the environment. These include: strontium-90, iodine-131 and plutonium-239. Strontium becomes concentrated in bones and iodine in the thyroid gland. Plutonium is itself a nuclear fuel and is also used in nuclear weapons. It is highly toxic. If breathed in, a very small amount can kill a person.

Atomic bomb

In an atomic bomb, an uncontrolled chain reaction takes place so that an extremely large amount of energy is released in a short interval of time. In the first atomic bomb shown in Figure 13.12 (a), an uncontrolled chain reaction was started by bringing two lumps of pure uranium-235 together so that the critical mass was exceeded. In present day nuclear weapons, plutonium-239 is used for fission.

When completely fissioned, 1 kg (2.2 pounds) of uranium-235 releases the energy equivalently produced by 17 000 tons, or 17 kilotons of trinitrotoluene (TNT). The detonation of an atomic bomb releases enormous amounts of thermal energy, or heat, achieving temperatures of several million degrees Celsius. This thermal energy creates a large fireball, the heat of which can ignite ground fires that can incinerate an entire small city. Convection currents created by the explosion suck dust and other ground materials up into the fireball, creating the characteristic mushroom-shaped cloud of an atomic explosion is shown in Figure 13.12 (b). The detonation also immediately produces a strong shock wave that propagates outward from the blast to distances of several miles, gradually losing its force along the way. Such a blast wave can destroy buildings for several miles from the location of the burst.



(a)

(b)

Figure 13.12 (a) The atomic bomb (Little Boy), dropped on Hiroshima
(b) the mushroom-shaped cloud of atomic explosion

3. **Environmental impact of nuclear energy**

A major environmental concern related to nuclear power is the creation of radioactive wastes such as spent (used) reactor fuel and other radioactive wastes (such as strontium-90, iodine-131 and cesium-137). These materials can remain radioactive and dangerous to human health for thousands of years. Spent nuclear fuel storage is mostly a problem.

When managed correctly, nuclear power plants are capable of being more environmentally-friendly than other traditional energy generation options. The greenhouse gas emissions from nuclear fission power are much smaller than those associated with coal, oil and gas, and the routine health risks are much smaller than those associated with coal.

However, there is a catastrophic risk potential if containment fails, which in nuclear reactors can be brought about by overheated fuels melting and releasing large quantities of fission products into the environment. This potential risk could wipe out the benefits.

The 1979 Three Mile Island accident and 1986 Chernobyl disaster ended the rapid growth of global nuclear power capacity. A release of radioactive materials followed the 2011 Japanese tsunami which damaged the Fukushima I Nuclear Power Plant, resulting in hydrogen gas explosions and partial meltdown. The large-scale release of radioactivity resulted in people being evacuated from a 20 km exclusion zone set up around the power plant, similar to the 30 km radius Chernobyl Exclusion Zone still in effect.

Reviewed Exercise

- What device is a nuclear reactor? Give a brief description of the essential component of a nuclear reactor.

Key Words: nuclear reaction, nuclear fission, chain reaction, nuclear reactor, radioactive waste, environmental impacts of nuclear energy

13.3 WAVE-PARTICLE DUALISM

In Grade 11, wave-particle duality for light has been learnt. In 1905, German physicist Albert Einstein proposed that light, considered a form of electromagnetic waves, must also be thought of as particle or photon.

In 1923, French physicist Louis de Broglie extended the idea of the wave-particle duality. He proposed that the wavelength of a material particle would be related to its momentum. Louis de Broglie suggested that the momentum p of a particle and its associated wavelength λ are related by the following equation.

$$\lambda = \frac{h}{p} \qquad (13.8)$$

The wavelength λ associated with a moving particle (matter wave) is also called de Broglie wavelength.

If a particle is moving with a velocity v, the momentum $p = mv$ and its wavelength $\lambda = \frac{h}{mv}$.

All moving particles exhibit properties of both particles and waves nature.

Momentum p relates kinetic energy KE of the particle as follows.

$$KE = \frac{1}{2} mv^2 = \frac{p^2}{2m}$$

$$p = \sqrt{2\,m\,KE}$$

The de Broglie wavelength of a particle in terms of its kinetic energy is expressed as,

$$\lambda = \frac{h}{\sqrt{2\,m\,KE}} \qquad (13.9)$$

The existence of matter waves were first experimentally confirmed by Thomson's cathode ray diffraction experiment and the Davisson-Germer experiment for electrons. The de Broglie hypothesis has also been confirmed for other elementary particles. Furthermore, neutral atoms and even molecules have been shown to be wave-like. In electron microscope, the wave nature of electrons is used instead of light.

Interpretation of Bohr's Second Postulate Using de Broglie Hypothesis

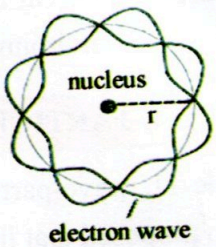
Bohr's second postulate of hydrogen atom can be explained with matter waves concept of electrons as follows.

An integral number of de Broglie wavelengths associated with a revolving electron must fit in the circumference of an orbit.

Therefore, $2\pi r = n\lambda$ where $n = 1,2,3,4\ldots$

The de Broglie wavelength of electron, $\lambda = \frac{h}{p}$

Thus, $2\pi r = \frac{nh}{p}$



If the speed of electron in an orbit is v, its momentum $p = mv$

$$2\pi r = \frac{nh}{mv}$$

$$mvr = n\frac{h}{2\pi}$$

$$mvr = n\hbar$$

Where $\hbar = \frac{h}{2\pi}$, h = Planck's constant, and mvr = angular momentum of the electron.

As a result, de Broglie hypothesis proves Bohr's second postulate, which states that the angular momentum of the orbiting electron is quantized.

Example 13.5 Calculate the de Broglie wavelength of (i) a 0.65 kg basketball thrown at a speed of 10 m s^{-1}, (ii) a nonrelativistic electron with a kinetic energy of 1 eV.

(i) $m = 0.65$ kg , $v = 10$ m s^{-1}, $h = 6.63 \times 10^{-34}$ J s

The de Broglie wavelength associated with basketball,

$$\lambda = \frac{h}{mv}$$

$$\lambda = \frac{6.63\times10^{-34}}{0.65\times10} = 1.02 \times 10^{-34} \text{ m}$$

(ii) $KE = 1$ eV $= 1.6\times 10^{-19}$ J

The de Broglie wavelength associated with electron,

$$\lambda = \frac{h}{\sqrt{2\,m\,KE}}$$

$$\lambda = \frac{6.63\times10^{-34}}{\sqrt{2\times9.1\times10^{-31}\times1.6\times10^{-19}}} = 1.23\times10^{-9}\ \text{m}$$

The de Broglie wavelength for basketball (macroscopic particle) is very much shorter than those of electron (microscopic particle). Hence, wave nature cannot be observed for basketball.

1 eV (electron volt) is the energy gained by an electron (a particle of charge e) when accelerated through a potential of one volt.

Reviewed Exercise

- How is the de Broglie wavelength of electrons related to the quantization of their orbits in atoms?

Key Words: de Broglie wavelength, matter wave, angular momentum

13.4 PRINCIPLE OF SPECIAL THEORY OF RELATIVITY

Frame of Reference and Speed of Light

We first have to learn some basic concepts and observations which lead to Einstein's development of his special theory of relativity. Firstly, we need to understand that motion is relative; and that observation and measurement of motion depends on the frame of reference. Galileo was the first to point out that all motion is relative. Motion is always measured with respect to a reference point (or place), which is called a frame of reference. The choice of the frame of reference is arbitrary. For example, the motion of a passenger in a moving bus can be viewed by two observers: (i) another passenger on the same bus (first observer), and (ii) a person standing on the sidewalk (second observer). For the first observer the passenger on the bus is at rest. On the other hand, for the second observer the passenger on the bus is moving with a velocity. Hence, frame of reference is an important concept in making observations and measurements.

Another important concept is the propagation and speed of light. During 19th century, it was generally believed that light could not travel across empty space. Most scientists believed that light, like all other types of energy that propagated as waves needed to travel through a material medium. It was considered that the space that light waves traveled through was filled with an invisible hypothetical material called ether (also spelled aether). So, it is expected that the speed of light measured by observers would depend on their relative motion with respect to ether.

In 1887, Albert Michelson and Edward Morley examined the existence ether by detecting the motion of the earth through ether. They designed an interferometer shown in Figure 13.13. A light beam is split into two perpendicular paths and then recombined. Recombining the light waves

produces an interference pattern of bright and dark fringes. If the earth is moving through ether a shift in the fringes of interference pattern would have observed because the speed of light in two perpendicular directions would be different.

However, no fringe shift is observed which leads to two important conclusions: that there is no ether and that the speed of light is the same regardless of the relative motion of source and observer. This result is generally considered to be the first strong evidence against the prevalent ether theory, as well as initiating a line of research that eventually led to the special theory of relativity.

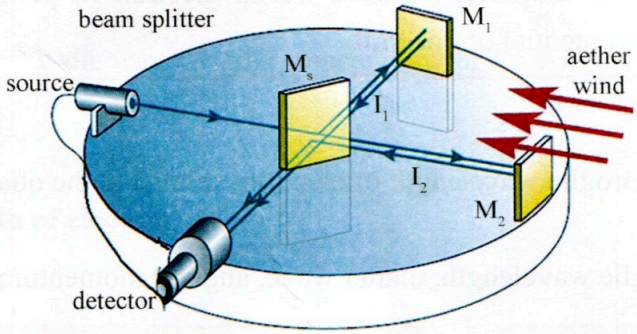


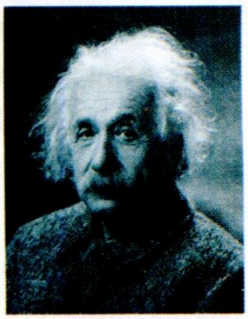
Figure 13.13 Michelson–Morley experiment

Einstein's Postulates

Einstein special theory of relativity is based on two simple postulates:

(i) The laws of physics are the same in all inertial frames of reference.

(ii) The speed of light in vacuum is the same in all inertial frames of reference and is not affected by the speed of its source.



Albert Einstein
(1879–1955)

An inertial frame is a frame of reference that has no acceleration (where all objects follow Newton's first law of motion). That is, all inertial frames are in a state of constant, rectilinear motion (constant velocity) with respect to one another. For example, the inside of a car moving along a road at constant velocity and the inside of a stationary house are inertial frames of reference.

The consequences of Einstein special theory of relativity are:

1. Relativity of simultaneity.
2. Relativistic effects on time interval (time dilation).
3. Relativistic effects on length (length contraction).
4. Relativistic effects on mass, momentum and energy.

1. **Relativity of simultaneity**

Relativity of simultaneity states that two events that are simultaneous in one frame of reference may not be simultaneous in another frame moving relative to the first frame. There are many thought experiments to explain this concept. To understand this idea, consider a thought experiment

Train-and-platform, suggested by Daniel Frost Comstock in 1910 and Einstein in 1917.
It consists of one observer midway inside a speeding train car and another observer standing on a platform as the train moves past. A flash of light is sent off at the centre of the train car just as the two observers pass each other.
For the observer on board the train (first observer), the front and back of the train car are at fixed distances from the light source. According to this observer, the light will reach the front and back of the train car at the same time (i.e. simultaneously) as illustrated in Figure 13.14(a).
However, for the observer standing on the platform (second observer), the back of the train car is moving toward the point at which the flash was sent off, and the front of the train car is moving away from it. Hence, the light headed towards the back of the train car will have less distance to cover than the light headed for the front. Since the speed of light is finite and the same in all directions, the flashes of light will strike the ends of the train car at different times for the second observer as illustrated in Figure 13.14(b).
Thus, for the first observer, the light will reach the front and back of the train car simultaneously; while for the second observer, this will not be the case. That is, if one reference frame assigns precisely the same time to two events that are at different points in space, another reference frame that is moving relative to the first will generally assign different times to the two events (the only exception being when motion is exactly perpendicular to the line connecting the locations of both events).

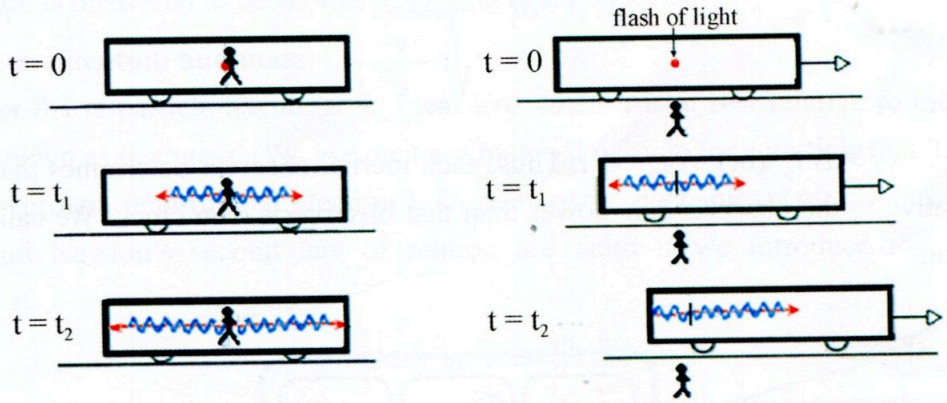


Figure 13.14 A thought experiment train and platform

2. **Relativistic effects on time interval (Time dilation)**

Time dilation is the phenomenon of time passing more slowly for an observer who is moving relative to another observer. That is, time dilation is the phenomenon of slowing down of time that occurs in a system in motion relative to an outside observer. This effect becomes apparent especially as the speed of the system approaches that of light.

Let us consider a thought experiment to explain time dilation. In Figure 13.15, an astronaut measures the time it takes for light to travel from the light source across the ship (space craft), bounce off a mirror, and return back. This time interval is obtained as Δt_0. The same time interval is measured by an earth-bound observer and obtained as Δt. It can be shown that time Δt_0 measured by the astronaut is smaller than the time Δt measured by the earth-bound observer.

The passage of time is different for the two observers because the distance 2 D, light travels in the astronaut's frame is smaller than the distance 2 S, light travels as observed by the earth-bound frame.

The time interval Δt_0 (known as the proper time) measured in the astronaut's frame is

$$\Delta t_0 = \frac{2D}{c}$$

The time interval Δt measured in the earth-bound observer's frame is

$$\Delta t = \frac{2S}{c}$$

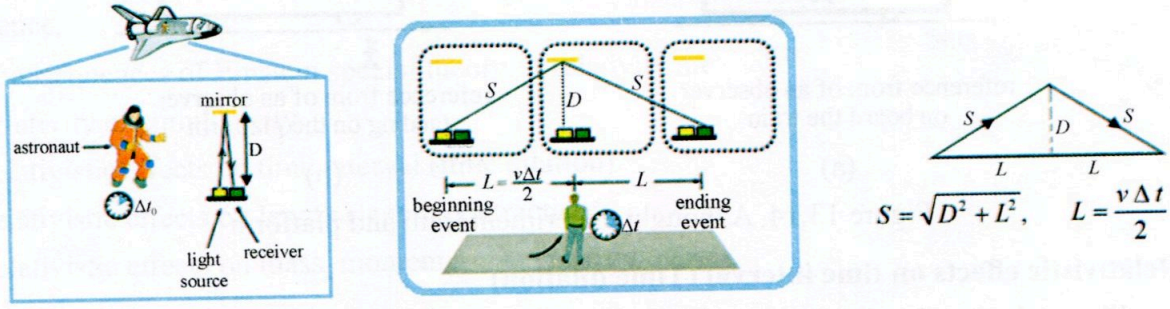
where $S = \sqrt{D^2 + L^2}$, and $L = \frac{v\,\Delta t}{2}$

Therefore, $$\Delta t = \frac{2}{c}\sqrt{(\frac{c\Delta t_0}{2})^2 + (\frac{v\Delta t}{2})^2}$$

squaring and solving for Δt, we obtain

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - \frac{v^2}{c^2}}} \qquad (13.10)$$

Since $v < c$, $\Delta t > \Delta t_0$ (or) $\Delta t_0 < \Delta t$. Thus, each inertial observer determines that all clocks in motion relative to the observer run slower than that observer's own clock. We call this effect as time dilation.



(a) (b) (c)

Figure 13.15 (a) The time Δt_0 measured by astronaut (b) the time Δt measured by earth-bound observer (c) relation between distances travelled by light

Equation (13.10) can be expressed as $\Delta t = \gamma \Delta t_0$ (13.11)

where the relativistic factor γ is $\gamma = \dfrac{1}{\sqrt{1 - \dfrac{v^2}{c^2}}}$

When the velocity v is small compared to the speed of light c, then v/c becomes small, and γ becomes close to 1. When this happens, time measurements are the same in both frames of reference. Relativistic effects, meaning those that have to do with special relativity, usually become significant when speeds become comparable to the speed of light.

3. **Relativistic effects on length (length contraction)**

Like the time interval between two events depends on the observer's frame of reference, the length between two points also depends on the observer's frame of reference. Let L_0 is the proper length of an object measured by an observer who is at rest relative to the object. The length of the same object measured by another observer who is moving with a constant velocity v in a direction parallel to the length of the object is L. We can also show that L_0 and L are related by the expression

$$L = \frac{L_0}{\gamma} = L_0 \sqrt{1 - \frac{v^2}{c^2}} \qquad (13.12)$$

Since $v < c$, $L < L_0$. That is, an object has a proper length L_0 when it is measured by an observer at rest with respect to the object. However, when the object is moving with speed v parallel to its length, its length is measured to be shorter according to Eq. (13.12).

4. **Relativistic momentum and mass**

Let us consider that a particle has mass m_0 measured when it is at rest relative to the observer. Hence, m_0 is known as rest mass. (We use the term material particle for a particle that has nonzero rest mass, like proton, neutron and electron.) In relativistic mechanics, law of conservation of momentum and Newton's second law of motion are valid if we introduce the relativistic momentum $\vec{p}$ as

$$\vec{p} = \gamma m_0 \vec{v} = \frac{m_0 \vec{v}}{\sqrt{1 - \dfrac{v^2}{c^2}}} \qquad (13.13)$$

where m_0 is the rest mass and $\vec{v}$ is the velocity of the particle as seen by a stationary observer. Eq. (13.13) for relativistic momentum is sometimes interpreted as a rapidly moving particle increases in mass. If a particle has mass m_0 when it is at rest, its relativistic mass m when moving with velocity $\vec{v}$ is

$$m = \frac{m_0}{\sqrt{1 - \dfrac{v^2}{c^2}}} \qquad (13.14)$$

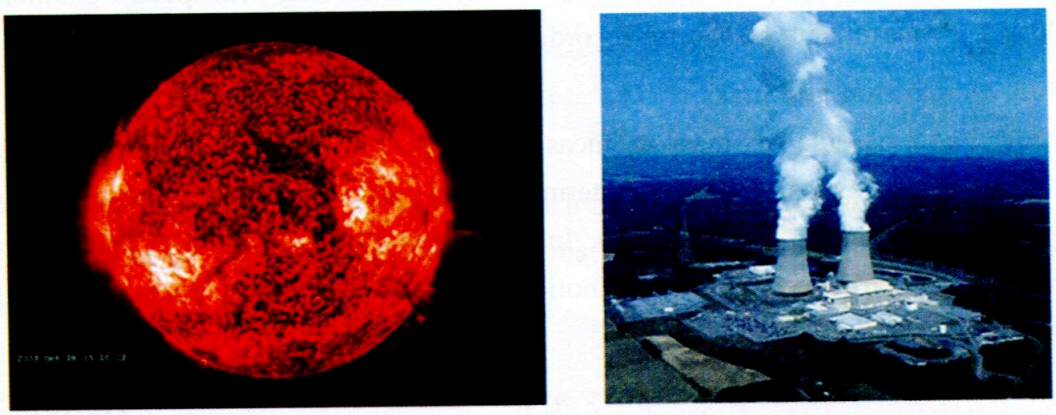
Relativistic Energy

Einstein showed that the law of conservation of energy is valid relativistically. A particle or object of rest mass m_0 moving at velocity v has relativistic energy given by

$$E = \gamma m_0 c^2 \qquad (13.15)$$

This is the expression for the total energy of an object of mass m_0 at speed v and includes both kinetic and potential energy. When an object is at rest, v is 0 and the relativistic factor γ is 1. Then the rest energy, m_0, is $E_0 = m_0 c^2$.

There are many experimental evidences that rest energy really exists. The energy released in nuclear reactions and decay of radioactive nuclei are good examples of conversion of mass into energy. The energy generated by nuclear reactions (fusion reaction) in the sun is radiated to earth in the form of electromagnetic radiation, where it is then transformed into all other forms and is used in our daily life. In nuclear power plants, the energy from nuclear fission reactions is an example of peaceful use of nuclear energy. In each of these cases, the source of the energy is the conversion of a small amount of mass into a large amount of energy. Figure 13.16 shows some source of energy released from conversion of mass into energy.



(a) (b)

Figure 13.16 Energy is released from conversion of mass into energy
(a) in the sun (b) in a nuclear reactor

Comparison between Einstein's special and general relativity theories

Table 10.1 Comparing special relativity and general relativity

Special Relativity	General Relativity
Published in 1905	Final form published in 1916
A theory of space-time	A theory of gravity
Applies to observers moving at constant speed (inertial frame)	Applies to observers that are accelerating (non-inertial frame)
Most useful in the field of nuclear physics	Most useful in the field of astrophysics
Accepted quickly and put to practical use by nuclear physicists and quantum chemists	Largely ignored until 1960 when new mathematical techniques made the theory more accessible and astronomers found some important applications

Also note that the theory of general relativity includes the theory of special relativity.

Example 13.6 A positron is a type of antimatter that is just like an electron, except that it has a positive charge. When a positron and an electron collide, their masses are completely annihilated and converted to energy in the form of gamma rays. If both particles have a rest mass of 9.11×10^{-31} kg, find the rest energy of each particle and the total energy of gamma rays in MeV.

$m_0 = 9.11 \times 10^{-31}$ kg, $c = 3 \times 10^8$ m s^{-1}

the rest energy of an electron or a positron is

$$E_0 = m_0 c^2$$
$$= 9.11 \times 10^{-31} \times (3 \times 10^8)^2 = 8.19 \times 10^{-14} \text{ J}$$

Since ($1 \text{ MeV} = 1.6 \times 10^{-13}$ J), $E_0 = \dfrac{8.19 \times 10^{-14}}{1.6 \times 10^{-13}} = 0.511$ MeV

the total energy of gamma rays is

$$E = 2 \times 0.511 = 1.022 \text{ MeV}$$

Example 13.7 One stary night, an earth-bound observer looking up the stars sees an extra-terrestrial spaceship flashacross the sky. The ship is 50 m long and is traveling at 95 percent of the speed of light. What would the ship's length be when measured from earth-bound frame of reference?

proper length of the ship, $L_0 = 50$ m; velocity, $v = 0.95\ c$

observed length L of the ship using relativistic length contraction,

$$L = \frac{L_0}{\gamma} = L_0 \sqrt{1 - \frac{v^2}{c^2}}$$
$$= 50 \sqrt{1 - \frac{(0.95\,c)^2}{c^2}}$$
$$= 50 \sqrt{1 - (0.95)^2} = 15.61 \text{ m}$$

Example 13.8 How fast must a vehicle travel for 1 s of time measured on a passenger's watch in the vehicle to differ by 1 % for an observer measuring it from the ground outside?

$\Delta t_0 = 1\ s$ and $\Delta t = 1.01\ s$ (differ by 1 %)

$$\frac{\Delta t_0}{\Delta t} = \frac{1}{1.01}$$

The time dilation is given by $\Delta t = \gamma \Delta t_0 = \dfrac{\Delta t_0}{\sqrt{1-\dfrac{v^2}{c^2}}}$

$$\frac{\Delta t_0}{\Delta t} = \sqrt{1-\frac{v^2}{c^2}}$$

$$(\frac{\Delta t_0}{\Delta t})^2 = 1-\frac{v^2}{c^2}$$

$$\frac{v^2}{c^2} = 1-(\frac{\Delta t_0}{\Delta t})^2$$

$$\frac{v}{c} = \sqrt{1-(\frac{\Delta t_0}{\Delta t})^2} = \sqrt{1-(\frac{1}{1.01})^2}$$

$$= 0.14$$

$$v = 0.14\ c$$

Reviewed Exercise

1. Calculate the relativistic factor γ, for a particle traveling at 99.7 percent of the speed of light.
2. Calculate the relativistic factor γ, for an object traveling at 2×10^8 m s^{-1}.

Key Words: frame of reference, special relativity, general relativity, simultaneity, mass defect, length contraction, time dilation, proper length, relativistic factor

SUMMARY

The radiations are emitted from the nucleus of uranium atoms and are of three types; namely alpha, beta and gamma rays. This phenomenon is known as **radioactivity**.

The **half-life** is defined as the time for half the atoms in a radioactive sample to decay.

A **nuclear reaction** is a process in which a particle or a gamma radiation penetrate into a nucleus and changes its configuration.

The wavelength λ associated with a moving particle (matter wave) is also called **de Broglie wavelength**.

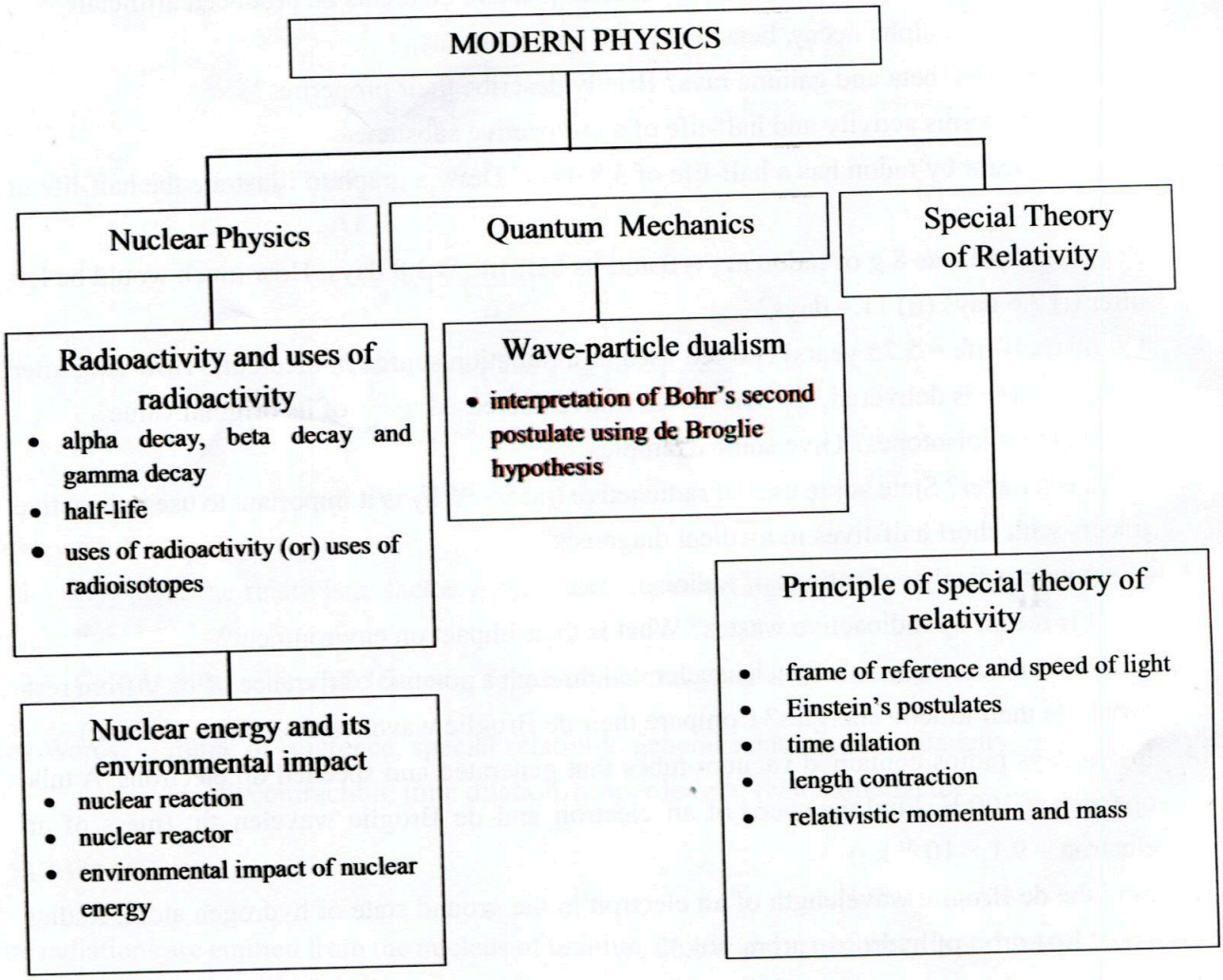
An **inertial frame** is a frame of reference that has no acceleration (where all objects follow Newton's first law of motion).

Time dilation is the phenomenon of time passing more slowly for an observer who is moving relative to another observer.

EXERCISES

1. What is radioactivity? Who discovered radioactivity? Name some radioactive elements.
2. Are all elements in nature radioactive? Can radioactive elements be produced artificially?
3. Briefly describe alpha decay, beta decay and gamma decay.
4. What are alpha, beta and gamma rays? Briefly describe their properties?
5. Explain the terms activity and half-life of a radioactive substance.
6. What is meant by radon has a half-life of 3.8 days? Draw a graph to illustrate the half-life of radon.
7. Suppose there was 8 g of radon at $t = 0$ and its half-life is 3.8 days. How much would be left after (i) 7.6 days (ii) 11.4 days?
8. Co-60 (half-life = 5.25 years) is often used as a radiation source in medicine. How long after a new sample is delivered, will the activity have decreased to $\frac{1}{8}$ of its original value?
9. What are radioisotopes? Give some examples.
10. What is a tracer? State some uses of radioactive tracers. Why is it important to use radioactive tracers with short half-lives in medical diagnosis?
11. Give two medical applications of radioisotopes.
12. What is meant by radioactive wastes? What is their impact on environment?
13. A proton and an electron are each accelerated through a potential difference of 1 kV from rest. What are their kinetic energies? Compare their de Broglie wavelengths. ($m_p = 1840\, m_e$)
14. In old days radios contained vacuum tubes that generated and speeded up electrons. A tube operates at 100 V. Find the speed of an electron and de Broglie wavelength. (mass of an electron = 9.1×10^{-31} kg)
15. Find the de Broglie wavelength of an electron in the ground state of hydrogen atom. Radius of the first orbit of hydrogen atom is 0.53 Å.
16. The distance between two points, called the proper length L_0 is 1 km. An observer in motion with respect to the frame of reference of the two points measures 0.8 km, which is L. What is the relative speed of the frame of reference with respect to the observer?

CONCEPT MAP


MODERN PHYSICS
Nuclear Physics
Quantum Mechanics
Special Theory of Relativity
Radioactivity and uses of radioactivity
• alpha decay, beta decay and gamma decay
• half-life
• uses of radioactivity (or) uses of radioisotopes
Wave-particle dualism
• interpretation of Bohr's second postulate using de Broglie hypothesis
Principle of special theory of relativity
• frame of reference and speed of light
• Einstein's postulates
• time dilation
• length contraction
• relativistic momentum and mass
Nuclear energy and its environmental impact
• nuclear reaction
• nuclear reactor
• environmental impact of nuclear energy

REFERENCES

Basic Education Curriculum, Syllabus and Textbook Committee, (2018-2019) *PHYSICS GRADE 10*. Yangon, Myanmar Times Publishing.

Basic Education Curriculum, Syllabus and Textbook Committee, (2018-2019) *PHYSICS GRADE 11*. Yangon, Myanmar Times Publishing.

Charles Chew and Chow Siew Foong, (1995) *Physics Matters*. 1st Edition, Singapore, Times Printers Ltd.

Charles Chew and Chow Siew Foong and Ho Boon Tiong, (2014) *Physics Matters, GCE 'O' Level*. 4th Edition, Singapore, Marshall Cavendish Education Pte Ltd.

Loo Wan Yong and Loo Kwok Wai, (2013) *All about Physics*. 2nd Edition, Singapore, Pearson Education South Asia Pte Ltd.

Sandra Mitchell and Chris Sherry, (2001) *PHYSICS for OCR A*. 1st Edition, Oxford, Reed Educational & Professional Publishing Ltd.

Tom Duncan and Heather Kennett, (2009) *IGCSE Physics*. 2nd Edition, Bembo, Fakenham Photosetting Ltd.

Loo Wan Yong and Loo Kwok Wai, (2007) *Physics Insights, 'O' Level*. 2nd Edition, Singapore, Pearson Education South Asia Pte Ltd.

Raymond A. Serway and John W. Jewett, (2013) *Principles of Physics, A Calculus-Based Text*. 5th Edition, Vol. 2, Canada, Nelson Education, Ltd.

Jerry D. Wilson, (1987) *Technical College Physics*. 2nd Edition, New York, Saunders College Publishing.

Raymond A. Serway and Jerry S. Faughn, (1989) *College Physics*. 2nd Edition, London, Saunders College Publishing.

Thomas L Floyd, (2006) *Digital Fundamentals*. 9th Edition, Pearson Education Ltd.